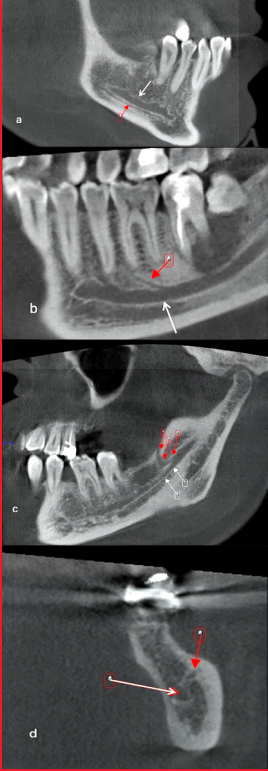




BEZMİÂLEM science



Comparative Efficacy of Diazepam Monotherapy Versus Diazepam-haloperidol Combination in the Treatment of Stimulant-induced Agitation: A Double-blind Clinical Trial

Faraz ZANDIYEH, Maral RAMEZANI, Shahin SHADNIA, Babak MOSTAFAZADEH, Mitra RAHIMI, Nava SOLEIMANI, Peyman ERFANTALABEVINI, Eka KURDADZE, Jemima D. Desir DOUGE; Tehran, Arak, Iran; Tbilisi, Georgia; Florida, USA

Measurement of Irisin, Adiponectin, and Leptin Levels in Patients with Idiopathic Epilepsy

Şeyma SÖNMEZ ŞAHİN, Türkan UYGUR ŞAHİN, Tuğçe YILDIZ, Zeynep Ebru ÇAKIN, Seher DURAN, Şahabettin SELEK, Dilara Füsün İÇAĞASIOĞLU, Akın IŞCAN; İstanbul, Siirt, Türkiye

Estimation of the Median Effective Dose of Propofol for Colonoscope Advancement to the Rectosigmoid Junction in Geriatric Patients: A Prospective Dose-finding Study

Bedirhan GÜNEL, Ayşe ŞENCAN; Kocaeli, Sakarya, Türkiye

Prognostic Value of Admission Red Cell Distribution Width and Mean Platelet Volume in Pediatric Sepsis and Septic Shock: Comparison with CRP, Procalcitonin, Lactate, PRISM IV and pSOFA

Ferhat SARI, Gökhan CEYLAN, Özlem SARAÇ, Gülhan ATAKUL, Ekin SOYDAN, Hasan AĞIN; İstanbul, İzmir, Aydın, Türkiye; Bonaduz, Switzerland

Volume 14 • Issue 4 • October 2026

bezmialemscience.org



Editor-in-Chief

Prof. Dr. Adem AKÇAKAYA

Bezmialem Vakıf University Faculty of Medicine, Department of General Surgery, İstanbul, Türkiye
E-mail: drakcakaya@gmail.com

Executive Deputy Chief Editor

Prof. Dr. Bülent DURDU

Bezmialem Vakıf University Faculty of Medicine, Department of Infectious Diseases and Clinical Microbiology, İstanbul, Türkiye
E-mail: bulentdurdu@gmail.com

Deputy Editors

Prof. Dr. Ali UZUNKÖY

Harran University Faculty of Medicine, Department of General Surgery, Şanlıurfa, Türkiye
E-mail: aliuzunkoy@yahoo.com

Prof. Dr. Remzi DOĞAN

Bezmialem Vakıf University Faculty of Medicine, Department of Otorhinolaryngology, İstanbul, Türkiye
E-mail: dr.remzidogan@gmail.com

Mahmut AKGÜL, MD

Brigham and Women's Hospital, Clinic of Pathology, Boston, Massachusetts, USA
E-mail: makgul@bwh.harvard.edu

Prof. Dr. Hayrettin DAŞKAYA

Bezmialem Vakıf University Faculty of Medicine, Department of Anesthesiology and Reanimation, İstanbul, Türkiye
E-mail: h.daskaya@gmail.com

Prof. Dr. İbrahim AYDOĞDU

Bezmialem Vakıf University Faculty of Medicine, Department of Pediatric Surgery, İstanbul, Türkiye
E-mail: draydogdu@yahoo.com

Associated Editors

Prof. Dr. Abdürrahim KOÇYİĞİT

Bezmialem University Faculty of Medicine, Department of Medical Biochemistry, İstanbul, Türkiye
E-mail: akocyigit@bezmialem.edu.tr

Prof. Dr. Simona CAVALU

University of Oradea, Faculty of Medicine and Pharmacy, Oradea, Romania
E-mail: simona.cavalu@gmail.com

Prof. Dr. Pinar SOYSAL

Bezmialem Vakıf University Faculty of Medicine, Department of Geriatric Medicine, İstanbul, Türkiye
E-mail: psoysal@bezmialem.edu.tr

Prof. Dr. Mehmet Burak GÜNEŞER

Bezmialem Vakıf University Faculty of Dentistry, Department of Endodontics, İstanbul, Türkiye
E-mail: mbguneser@bezmialem.edu.tr

Assoc. Prof. Büşra Yaprak BAYRAK

Kocaeli University Faculty of Medicine, Department of Pathology, Kocaeli, Türkiye
E-mail: busra.yaprakbayrak@kocaeli.edu.tr

Section Editors

Assoc. Prof. Alis KOSTANOĞLU

Bezmialem Vakıf University Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, İstanbul, Türkiye
E-mail: aliskostanoglu@gmail.com

Assoc. Prof. Ayşe Filiz GÖKMEN KARASU

Bezmialem Vakıf University Faculty of Medicine, Department of Gynecology and Obstetrics, İstanbul, Türkiye
E-mail: afgangmen@gmail.com

Assoc. Prof. Ayşegül DOĞAN DEMİR

İstanbul Medipol University Faculty of Medicine, Department of Social Pediatrics, İstanbul, Türkiye
E-mail: ayseguldogandemir@gmail.com

Assoc. Prof. Bahadır TASLIDERE

Bezmialem Vakıf University Faculty of Medicine, Department of Emergency Medicine, İstanbul, Türkiye
E-mail: btaslidere@bezmialem.edu.tr

Assoc. Prof. Burcu OĞLAĞI ÖZKOÇ

Bezmialem Vakıf University Faculty of Dentistry, Department of Restorative Dental Treatment, İstanbul, Türkiye
E-mail: burcu923@hotmail.com

Assoc. Prof. Çağla KIZILARSLAN HANÇER

Bezmialem University Faculty of Pharmacy, Department of Pharmaceutical Botany, İstanbul, Türkiye
E-mail: c.kizilarслан@gmail.com

Assoc. Prof. Ebru HACIOSMANOĞLU ALDOĞAN

İstanbul-Cerrahpaşa University Faculty of Medicine, Department of Biophysics, İstanbul, Türkiye
E-mail: ebru.aldoğan@iuc.edu.tr

Prof. Dr. Fahri AKBAŞ

Bezmialem Vakıf University Faculty of Medicine, Department of Medical Biology, İstanbul, Türkiye
E-mail: fakbas@bezmialem.edu.tr

Fatma Betül AKÇAKAYA ÖZER, MD

Republic of Türkiye Ministry of Health, Kağıthane District Health Directorate, İstanbul, Türkiye
E-mail: betulakcakaya@hotmail.com

Assoc. Prof. Mehmet Ali GÜLTEKİN

Bezmialem Vakıf University Faculty of Medicine, Department of Radiology, İstanbul, Türkiye
E-mail: mgultekin@bezmialem.edu.tr

Lect. Merve MEŞEDÜZÜ

Bezmialem Vakıf University Faculty of Health Sciences, Department of Nursing, İstanbul, Türkiye
E-mail: mmeseduzu@bezmialem.edu.tr

Assoc. Prof. Muhammed Batuhan AYIK

Bezmialem Vakıf University Faculty of Medicine, Department of Psychiatry, İstanbul, Türkiye
E-mail: batuhan.ayik@bezmialem.edu.tr

Prof. Dr. Muharrem KISKAÇ

Prof. Dr. Muharrem Kıskaç Private Clinic, Department of Internal Medicine, İstanbul, Türkiye
E-mail: muharremkiskac@gmail.com

Prof. Dr. Özlem SU KÜÇÜK

Bezmialem Vakıf University Faculty of Medicine, Department of Dermatology, İstanbul, Türkiye
E-mail: ozlemsukucuk2@yahoo.com.tr

Assoc. Prof. Semiramis ÖZYILMAZ

Bezmialem University Faculty of Medicine, Department of Physiotherapy and Rehabilitation, İstanbul, Türkiye
E-mail: sozyilmaz@bezmialem.edu.tr

Prof. Dr. Yeter DEMİR USLU

İstanbul Medipol University Faculty of Health Sciences, Department of Health Management, İstanbul, Türkiye
E-mail: yuslu@medipol.edu.tr

Assoc. Prof. Ozan Volkan YURDAKUL

Bezmialem Vakıf University Faculty of Medicine, Department of Physical Medicine and Rehabilitation, İstanbul, Türkiye
E-mail: oyurdakul@bezmialem.edu.tr

Biostatistics Consultants

Asst. Prof. Ömer UYSAL

Bezmialem Vakıf University Faculty of Medicine, Department of Biostatistics and Medical Informatics, Division of Basic Medical Sciences, İstanbul, Türkiye
E-mail: omer.uyasal@iuc.edu.tr

Assoc. Prof. Ayşegül YABACI TAK

Bezmialem Vakıf University Faculty of Medicine, Department of Biostatistics and Medical Informatics, İstanbul, Türkiye
E-mail: ayabaci@bezmialem.edu.tr

Scientific Board

Anne-Catherine ANDRES

Department of Clinical Research, University of Bern School of Medicine, Switzerland

Amrita BANERJEE

Department of Pharmaceutical Sciences, North Dakota State University School of Pharmacy, Fargo, ND, USA
ORCID: 0000-0002-1977-4685

Gökçen BAŞARANOĞLU

Department of Anesthesiology and Reanimation, Bezmialem Vakıf University School of Medicine, İstanbul, Türkiye

Artur BEKE

Department of Obstetrics and Gynecology, Semmelweis University, Budapest, Hungary

Ahmet BELCE

Department of Biochemistry, Biruni University School of Medicine, İstanbul, Türkiye
ORCID: 0000-0002-1228-7999

Ufuk ÇAKATAY

Department of Biochemistry, İstanbul University-Cerrahpaşa, Cerrahpaşa School of Medicine, İstanbul, Türkiye

Yeter DEMİR

Department of Healthcare Management, İstanbul Medipol University Faculty of Health Sciences, İstanbul, Türkiye

İrem Yağmur DİKER

Department of Molecular Biology and Genetics, Bezmialem Vakıf University, İstanbul, Türkiye

Özlem DURMAZ

Department of Pediatric Gastroenterology, Hepatology and Nutrition, İstanbul University School of Medicine, İstanbul, Türkiye

Atilla EROĞLU

Department of Thoracic Surgery, Atatürk University School of Medicine, Erzurum, Türkiye

Mukkades EŞREFOĞLU

Department of Histology and Embryology, Bezmialem Vakıf University, İstanbul, Türkiye

Joachim FANDREY

Department of Physiology, Duisburg University School of Medicine, Duisburg, Germany
ORCID: 0000-0001-9585-0531

Max GASSMAN

Department of Veterinary Physiology, Institute of Veterinary Physiology, University of Zurich, Zurich, Switzerland

Renate GAY

Department of Rheumatology, University of Zurich School of Medicine, Zurich, Switzerland

Steffen GAY

Department of Rheumatology, University of Zurich School of Medicine, Zurich, Switzerland

Klaus W. GRAETZ

Department of Cranio-Maxillo-Facial and Oral Surgery, Zurich University School of Medicine, Zurich, Switzerland

Ülkan KILIÇ

Department of Medical Biology, University of Health Sciences Türkiye School of Medicine, İstanbul, Türkiye

İsmet KIRPINAR

Private Practitioner, İstanbul, Türkiye

Abdürrahim KOÇYİĞİT

Department of Medical Biochemistry, Bezmialem Vakıf University, İstanbul, Türkiye

Tufan KUTLU

Department of Pediatric Gastroenterology and Hepatology, İstanbul University-Cerrahpaşa, Cerrahpaşa School of Medicine, İstanbul, Türkiye

Thomas A. LUTZ

Department of Veterinary Physiology, University of Zürich School of Medicine, Zurich, Switzerland

Martina MUCKENTHALER

Department of Pediatric Oncology, Heidelberg University, Heidelberg, Germany
ORCID: 0000-0002-3778-510X

Hayat ÖNYÜKSEL

Department of Biopharmaceutical Sciences, UIC Faculty of Pharmacy, Illinois, USA

Orhan ÖZTURAN

Department of Otolaryngology, Bezmialem Vakıf University School of Medicine, İstanbul, Türkiye

Şahabettin SELEK

Department of Medical Biochemistry, Bezmialem Vakıf University School of Medicine, İstanbul, Türkiye

Suhair SUNOQROT

Department of Pharmacy, Al-Zaytoonah University of Jordan
School of Pharmacy, Amman, Jordan

Claudiu T. SUPURAN

Department of Neuropharma, University of Florence School of
Medicine, Firenze, Italy

ORCID: 0000-0003-4262-0323

Arzu TEZVERGİL MUTLUAY

Department of Prosthetic, University of Turku School of
Medicine, Turku, Finland

ORCID: 0000-0003-0932-8531

Gülaçtı TOPÇU

Dean of the Faculty of Pharmacy, Bezmialem Vakıf University,
İstanbul, Türkiye

Oliver ULRICH

Department of Anatomy, University of Zurich School of
Medicine, Zurich, Switzerland

Yener YÖRÜK

Department of Thoracic Surgery, Trakya University School of
Medicine, Edirne, Türkiye

Sevgi CANBAZ

Department of Public Health, İstanbul University, İstanbul
Faculty of Medicine, İstanbul, Türkiye

Jie ZHOU

Department of Anesthesiology, Perioperative and Pain
Medicine, Brigham and Women's Hospital, Harvard Medical
School, Boston, MA, USA

Please refer to the journal's webpage (<https://www.bezmialemscience.org/>) for "Aims and Scope", "Instructions to Authors" and "Ethical Policy".

The editorial and publication process of Bezmialem Science are shaped in accordance with the guidelines of ICMJE, WAME, CSE, COPE, EASE, and NISO. The journal is in conformity with the Principles of Transparency and Best Practice in Scholarly Publishing.

Bezmialem Science is indexed in **Web of Science-Emerging Sources Citation Index, TUBITAK ULAKBİM, EBSCO, Gale, Embase, CABI, ProQuest, CINAHL, Türk Medline, Türk Atıf Dizini, İdealOnline, J-Gate, DOAJ, Hinari, GOALI, ARDI, OARE, AGORA** and CNKI.

The journal is published electronically.

Owner: Bezmialem Vakıf University

Responsible Manager: Adem AKÇAKAYA

CONTENTS

Commentary

- 321 Methodological Limitations in Global Burden of Disease Alcohol Models: A Hepatology-oriented Perspective
Atilla BEKTAŞ, Meltem ULUSOY, Tayfun UZBAY; Ankara, İstanbul, Türkiye
-

Original Articles

- 326 Comparative Efficacy of Diazepam Monotherapy Versus Diazepam-haloperidol Combination in the Treatment of Stimulant-induced Agitation: A Double-blind Clinical Trial
Faraz ZANDIYEH, Maral RAMEZANI, Shahin SHADNIA, Babak MOSTAFAZADEH, Mitra RAHIMI, Nava SOLEIMANI, Peyman ERFANTALABEVINI, Eka KURDADZE, Jemima D. Desir DOUGE; Tehran, Arak, Iran; Tbilisi, Georgia; Florida, USA
-
- 333 The Relationship Between Body Image Satisfaction and Depressive Symptoms in Living Liver Donors
Tuğba Nur ÖDEN, Tufan GÜMÜŞ, Şirin DURMUŞ, Alper UĞUZ, Murat ZEYTÜNLÜ; İzmir, Türkiye
-
- 341 The Effect of Standardized Patient Simulation on Nursing Students' Insulin Administration Skills
Hilal TÜZER, Bahar VARDAR İNKAYA, Şenay SARMASOĞLU KILIKÇIER, Melih ELÇİN, Betül ŞAHİN KILINÇ; Ankara, Bursa, Türkiye; Springfield, Massachusetts, USA
-
- 349 Evaluation of the Prevalence of Mandibular Canal Variations Using Cone-beam Computed Tomography
Elifhan ALAGÖZ, Nilüfer KARAÇAY, Nilüfer GÜRSOY, İrfan SARICA; İstanbul, Türkiye
-
- 357 Establishment of Age-specific Reference Intervals for Serum Magnesium Concentrations in Healthy Children Aged 0-18 Years: A Comparative Analysis of Non-parametric, Bootstrap, and Gaussian Mixture Model Approaches
Adile ÖZERTÜRK, Fatma KUMBASAR, Mehmet Murat YEKREK, Özge PASİN, Eray Metin GÜLER; İstanbul, Türkiye
-
- 365 Measurement of Irisin, Adiponectin, and Leptin Levels in Patients with Idiopathic Epilepsy
Şeyma SÖNMEZ ŞAHİN, Türkan UYGUR ŞAHİN, Tuğçe YILDIZ, Zeynep Ebru ÇAKIN, Seher DURAN, Şahabettin SELEK, Dilara Fusun İÇAĞASIOĞLU, Akın İŞCAN; İstanbul, Siirt, Türkiye
-
- 376 Comparative Analysis of Aesthetic Judgment between Orthodontists, General Dentists, and Different Versions of Large Language Models Using the IOTN-AC Index
Rümeysa Cemre ÖNCÜ, Aylin PAŞAOĞLU BOZKURT; İstanbul, Türkiye
-

CONTENTS

385 AHEAD Score and Prolonged Intensive Care Unit Stay Following Isolated Coronary Artery Bypass Grafting

Büşra GÜVENDİ ŞENGÖR, Mehmet ŞENGÖR, Muhammet Mucahit TİRYAKİ, Faruk AKALIN, Kadir BIYIKLI, Mustafa Ferhat KETEN, Cemalettin YILMAZ, Ünal AYDIN, Regayip ZEHİR; İstanbul, Yalova, Türkiye

393 Estimation of the Median Effective Dose of Propofol for Colonoscope Advancement to the Rectosigmoid Junction in Geriatric Patients: A Prospective Dose-finding Study

Bedirhan GÜNEL, Ayşe ŞENCAN; Kocaeli, Sakarya, Türkiye

401 Prognostic Value of Admission Red Cell Distribution Width and Mean Platelet Volume in Pediatric Sepsis and Septic Shock: Comparison with CRP, Procalcitonin, Lactate, PRISM IV and pSOFA

Ferhat SARI, Gökhan CEYLAN, Özlem SARAÇ, Gülhan ATAKUL, Ekin SOYDAN, Hasan AĞIN; İstanbul, İzmir, Aydın, Türkiye; Bonaduz, Switzerland

Case Report

410 Radiologically Presumed Needle-tract Seeding After Radiofrequency Ablation for Hepatocellular Carcinoma: A Case Report

Hacer KUNDAKCIOĞLU, Hüseyin TOPRAK, Adem AKÇAKAYA; Kahramanmaraş, İstanbul, Türkiye

Review

415 Physiotherapy for Pelvic Dysfunction: A Comprehensive Review

Özge BAYKAN ÇOPUROĞLU, Betül ERGÜN; Kayseri, Elazığ, Türkiye

Index

2026 Reviewer Index

EDITORIAL

Dear Readers,

We are delighted to present the final issue of Bezmialem Science for 2026 to our esteemed readers. The ever-evolving nature of the health sciences necessitates the sharing of new research findings with the scientific community, as well as the re-evaluation of existing knowledge in light of contemporary approaches. With this perspective, our journal has continued throughout the year to introduce original, high-quality, and innovative studies from various disciplines within the health sciences to the scientific community, while fostering scholarly interaction among researchers and contributing to the advancement of academic literature. Inspired by the deep-rooted foundation tradition of Bezmialem Vakıf University, we consider it a core tenet of our publishing philosophy to integrate scientific output with a human-centered approach and to support the potential of research to enhance healthcare services.

The cover image for this issue is drawn from the study titled “Evaluation of the Prevalence of Mandibular Canal Variations Using Cone-beam Computed Tomography,” authored by Alagöz E, Karaçay N, Gürsoy N, and Sarıca İ. In this study, which examines the prevalence and distribution of mandibular canal variations using cone-beam computed tomography (CBCT) images, 794 hemimandibular scans from 426 patients were analyzed to investigate the relationships between these variations and factors such as age, gender, edentulism, and side. The findings reveal that mandibular canal variations are quite common, with the retromolar canal being the most frequently observed variation. Furthermore, the study identifies a significant association between these variations and both age and edentulous status. The study highlights the importance of a detailed assessment of mandibular canal anatomy using CBCT prior to invasive dental procedures—such as implant placement and third molar surgery—in order to prevent potential neurovascular injuries and surgical complications.

Other articles featured in this issue include:

“Comparative Efficacy of Diazepam Monotherapy Versus Diazepam-haloperidol Combination in the Treatment of Stimulant-induced Agitation: A Double-blind Clinical Trial” by Zandiyeh F et al.

“Measurement of Irisin, Adiponectin, and Leptin Levels in Patients with Idiopathic Epilepsy” by Sönmez Şahin Ş et al.

“Estimation of the Median Effective Dose of Propofol for Colonoscopy Advancement to the Rectosigmoid Junction in Geriatric Patients: A Prospective Dose-finding Study” by Günel B and Şencan A.

“Prognostic Value of Admission Red Cell Distribution Width and Mean Platelet Volume in Pediatric Sepsis and Septic Shock: Comparison with CRP, Procalcitonin, Lactate, PRISM IV and pSOFA” by Sarı F et al.

In today's world, scientific progress gains meaning not merely through the generation of new knowledge, but through the critical evaluation of such information, its exchange across diverse disciplines, and its translation into practice. Furthermore, the quality of scientific publishing is fundamentally defined by the reliability of research, the rigor of the peer-review process, and the contribution that the generated knowledge makes to the scientific community. In this spirit, Bezmialem Science remains committed to providing a platform for research conducted in accordance with scientific methods and ethical principles, and to integrating scholarly contributions from various fields of expertise. We continue to work steadfastly toward enhancing our journal's academic standards and strengthening its scientific visibility—both nationally and internationally.

As we reach the final issue of the year, I would like to extend my sincere gratitude to all the authors who have contributed to the scientific development of our journal, as well as to our editors and reviewers, who have carried out the evaluation process with such diligence. The sustainability of scientific publishing is made possible through the collective effort and dedication of all stakeholders, each assuming their respective responsibilities. I believe that every valuable contribution made throughout this process plays a significant role in maintaining and enhancing the publication quality of our journal. Furthermore, I would like to thank all our readers—those who follow our journal, benefit from the published research, and support the advancement of the academic community through their own contributions to scientific knowledge.

As we look toward the year 2027, our goal is for Bezmialem Science to continue contributing to the field of health sciences by upholding a publishing philosophy rooted in scientific rigor, ethical standards, and interdisciplinary engagement. I hope the new year brings our researchers fruitful collaborations, original ideas, and high-quality scientific work; I wish all our readers a healthy, successful, and productive year ahead. Thank you for your interest and support throughout the year, and I look forward to connecting with you again in our upcoming issues.

Happy reading.

Prof. Dr. Adem AKÇAKAYA

Editor-in-Chief

Methodological Limitations in Global Burden of Disease Alcohol Models: A Hepatology-oriented Perspective

Global Hastalık Yüğü Alkol Modellerinde Metodolojik Sınırlılıklar: Hepatoloji Odaklı Bir Değerlendirme

 Atilla BEKTAŞ¹,  Meltem ULUSOY²,  Tayfun UZBAY³

¹Private Ankara Surgery Medical Center, Clinic of Gastroenterology, Ankara, Türkiye

²Hacettepe University Faculty of Science, Department of Biology, Division of Applied Biology, Ankara, Türkiye

³Üsküdar University Faculty of Medicine, Neuropsychopharmacology Application and Research Center, İstanbul, Türkiye

ABSTRACT

To critically evaluate the methodological limitations of global burden of disease (GBD) alcohol models from a hepatology-oriented perspective and to highlight the inadequate representation of liver-specific outcomes in current risk frameworks. This review analyzes the GBD alcohol studies from 2016, 2018, and 2022, examining their interpretive inconsistencies, methodological approaches, and the underrepresentation of alcohol-attributable liver diseases. Key concepts such as theoretical minimum risk exposure level (TMREL) and non-drinker equivalence (NDE) are critically assessed for their clinical applicability in hepatology. Despite relying on similar datasets, GBD studies have presented conflicting conclusions regarding safe alcohol consumption thresholds. The 2016 analysis suggested cardiovascular benefits from low-to-moderate intake, while the 2018 report—without new data—concluded all alcohol levels are harmful. The 2022 models introduced TMREL and NDE but failed to adequately account for hepatotoxic effects such as alcoholic hepatitis, cirrhosis, and hepatocellular carcinoma. Current frameworks overlook beverage type differences, consumption patterns, and comorbid liver conditions such as metabolic dysfunction-associated steatotic liver disease. Existing GBD alcohol models have significant methodological shortcomings that compromise their clinical applicability in hepatology. Future risk frameworks must integrate liver-centered endpoints, differentiate between beverage types,

Öz

Hepatoloji odaklı bir bakış açısıyla küresel hastalık yükü (GBD) alkol modellerinin metodolojik sınırlılıklarını eleştirel bir şekilde değerlendirmek ve mevcut risk çerçevelerinde karaciğere özgü sonuçların yetersiz temsilini vurgulamak. Bu derleme, 2016, 2018 ve 2022 yıllarına ait GBD alkol çalışmalarını analiz ederek, yorumlamadaki tutarsızlıkları, metodolojik yaklaşımları ve alkole bağlı karaciğer hastalıklarının yetersiz temsilini incelemektedir. Teorik minimum risk maruziyet düzeyi (TMREL) ve alkol kullanmayan eşdeğeri (NDE) gibi temel kavramlar, hepatolojideki klinik uygulanabilirlikleri açısından eleştirel olarak değerlendirilmektedir. Benzer veri kümelerine dayanmalarına rağmen, GBD çalışmaları güvenli alkol tüketim eşikleri konusunda çelişkili sonuçlar ortaya koymuştur. 2016 analizi, düşük ila orta düzeyde alımın kardiyovasküler faydalar sağladığını öne sürerken, 2018 raporu—yeni veriler olmaksızın—tüm alkol seviyelerinin zararlı olduğu sonucuna varmıştır. 2022 modelleri TMREL ve NDE'yi tanıtmıştır ancak alkolik hepatit, siroz ve hepatoselüler karsinom gibi hepatotoksik etkileri yeterince hesaba katmamıştır. Mevcut çerçeveler, içecek türü farklılıklarını, tüketim modellerini ve metabolik disfonksiyonla ilişkili steatotik karaciğer hastalığı gibi eşlik eden karaciğer rahatsızlıklarını göz ardı etmektedir. Mevcut GBD alkol modelleri, hepatolojideki klinik uygulanabilirliklerini tehlikeye atan önemli metodolojik eksikliklere sahiptir. Gelecekteki risk çerçeveleri, karaciğer merkezli sonlanım noktalarını entegre etmeli, içecek türleri

Address for Correspondence: Atilla Bektaş, MD, Private Ankara Surgery Medical Center, Clinic of Gastroenterology, Ankara, Türkiye

E-mail: atillabektas2000@yahoo.com

ORCID IDs of the authors: A.B.: 0000-0001-7775-3770, M.U.: 0000-0002-3885-2714, T.U.: 0000-0002-9784-5637

Cite this article as: Bektaş A, Ulusoy M, Uzbay T. Methodological limitations in global burden of disease alcohol models: a hepatology-oriented perspective. Bezmialem Science. 2026;14(4):321-325



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

Received: 19.07.2025

Accepted: 12.11.2025

Published date: 26.10.2026

ABSTRACT

and stratify risk by comorbidities. Evidence-based guidelines anchored in randomized controlled trials, systematic reviews, and meta-analyses are essential to provide clinically meaningful and organ-specific public health recommendations.

Keywords: Alcohol, liver cirrhosis, global burden of disease, risk modeling, MASLD

Öz

arasında ayırım yapmalı ve riski eşlik eden hastalıklara göre sınıflandırmalıdır. Klinik olarak anlamlı ve organa özgü halk sağlığı önerileri sunmak için randomize kontrollü çalışmalar, sistematik incelemeler ve meta-analizlere dayanan kanıta dayalı kılavuzlar şarttır.

Anahtar Kelimeler: Alkol, karaciğer sirozu, küresel hastalık yükü, risk modellemesi, MASLD

Introduction

The global burden of disease (GBD) alcohol studies from 2016, 2018, and 2022, though based on overlapping datasets, have generated contradictory conclusions. The 2016 analysis suggested a modest cardioprotective effect of low-to-moderate alcohol consumption, while the 2018 editorial—without new data—claimed that no amount of alcohol is safe, and the 2022 update introduced theoretical minimum risk exposure level (TMREL) and non-drinker equivalence (NDE) metrics recommending zero alcohol intake for those under 40, but permitting up to 1.87 drinks per day for older adults (1-4). These shifting thresholds illustrate the interpretive instability of observational models and the need for greater methodological consistency in population-level alcohol research.

Importantly, these models have consistently underrepresented liver-specific outcomes. Despite the liver being the primary site of alcohol-induced toxicity, GBD frameworks have prioritized cardiovascular metrics, neglecting alcoholic hepatitis, cirrhosis, and hepatocellular carcinoma—conditions with high global burden and direct mechanistic links to alcohol metabolism (1,4). Furthermore, GBD models fail to distinguish between beverage types, overlooking the metabolic and toxicological variability across beer, wine, and distilled spirits, including differential effects on oxidative stress and hepatic inflammation (5,6).

Notably, this underrepresentation persists despite liver-related mortality accounting for a significant and growing proportion of alcohol-attributable deaths, particularly in regions with rising [metabolic dysfunction-associated steatotic liver disease (MASLD); previously known as non-alcoholic fatty liver disease (NAFLD)] prevalence and region-specific liver mortality trends (7,8).

Future alcohol policy models must integrate liver-centered endpoints and stratify risk by alcohol type, dose, and comorbidities such as MASLD. Transparent, evidence-based frameworks—grounded in randomized trials and meta-analyses—are essential to improve the clinical applicability and public credibility of future guidelines (9,10). These concerns echo the hepatology-centered orientation of recent discourse in the Journal of Hepatology and underscore the importance of specialty-driven modeling. Clearer communication of age-specific and organ-specific risks will enhance hepatology’s role in shaping alcohol-related health policy (Table 1) (1,11).

Interpretive Inconsistencies and Methodological Limitations

The GBD 2016 study, a comprehensive systematic analysis covering 195 countries, concluded that low-to-moderate alcohol consumption (e.g., 1-2 standard drinks per day) could provide some cardiovascular benefits (2). However, the 2018 editorial commentary, which did not introduce

Table 1. Comparative summary of GBD alcohol modeling studies (2016-2022) and hepatology-focused critique

Study	Design	Methodology	Main findings	Limitations	Reference
GBD 2016	Systematic analysis (195 countries)	Observational cohort data	Low-to-moderate alcohol intake may have cardioprotective effects	Residual confounding, underreporting, no liver-specific stratification	(2)
GBD 2018	Editorial commentary	Narrative reinterpretation of 2016 data	No safe level of alcohol consumption	No new data, interpretative overreach, inconsistency with prior GBD	(3)
GBD 2022	Statistical modeling update	Statistical modeling using TMREL and NDE metrics by demographic strata	Risk varies by demographic group; zero alcohol recommended for <40 years	Complex metrics, limited clinical applicability, ambiguous guidance	(4)
Musto et al. (1)	Expert commentary (hepatology-oriented)	Critical review of GBD frameworks	Oversimplified messaging; lack of liver-specific indicators; need for hepatology-centered risk models	Highlights gap between population data and liver disease burden	(1)

GBD: Global burden of disease, TMREL: Theoretical minimum risk exposure level, NDE: Non-drinker equivalence

any new statistical analysis, contradicted these findings by claiming that all levels of alcohol consumption are harmful (3). The 2022 publication recalculated alcohol risks based on age, sex, and geographical differences, introducing TMREL and NDE metrics (4). These concepts, though statistically nuanced, are difficult to apply in practice and present challenges for policy development. The TMREL was determined to be zero alcohol for individuals under 40 years of age, while being acceptable at a certain level—up to 1.87 standard drinks—for those over 40. This age-based distinction has caused confusion in public health messaging and limits its applicability in clinical guidelines.

The GBD 2016 study's relative risk curve for all-cause outcomes showed a modest protective effect for low-level alcohol consumption, especially in cardiovascular diseases (2). However, the 2018 commentary dismissed this interpretation and emphasized a zero-risk stance (3). This sharp contrast demonstrates how observational data can be interpreted flexibly and, at times, inconsistently. The risk differential between abstainers and those consuming one drink per day is extremely small—and arguably not clinically significant—yet it formed the basis for a shift in global guidance (2-4). Such variability in interpretation underscores a broader methodological fragility in observational studies. These analyses are prone to biases, including underreporting, social desirability effects, and failure to account for binge drinking or irregular drinking patterns (9,10). For high-stakes decisions in health policy, reliance on observational data alone is insufficient; more robust methodologies such as randomized controlled trials (RCTs) and meta-analyses should be emphasized in future evaluations (10).

The Case for Liver-specific Risk Modeling

One of the fundamental methodological shortcomings identified by Musto et al. (1) is the inadequate representation of alcohol-attributable liver diseases in GBD alcohol models.

Hepatic complications such as alcoholic hepatitis, cirrhosis, and hepatocellular carcinoma, which are associated with high morbidity and mortality, constitute a significant proportion of the global alcohol-related disease burden. The direct toxic processes initiated by alcohol in the liver should be assessed before systemic effects, and the liver must be recognized as the primary target organ of alcohol toxicity (1,7,8).

Therefore, future public health policies and risk prediction models should prioritize liver-specific parameters. Although hepatic consequences—including alcoholic steatohepatitis, cirrhosis, and hepatocellular carcinoma—are major contributors to morbidity and mortality globally, they receive disproportionately less attention compared to cardiovascular endpoints. In hepatology and gastroenterology practice, hepatic effects are at least as critical as cardiovascular ones and must be adequately

represented in population-level modeling. Moreover, subclinical hepatic inflammation and fibrosis progression may go undetected in individuals consuming “moderate” amounts, especially when risk thresholds are generalized across populations (1,2,4).

Additionally, the term “alcohol” in GBD reports is typically applied generically, neglecting key differences among beverage types. Ethanol-containing drinks such as beer, wine, and distilled spirits differ significantly in polyphenol content, metabolic profiles, and associated toxicities (5). The cardioprotective properties of red wine, attributed to polyphenols such as resveratrol, have been acknowledged in the context of the French paradox (12), and moderate red-wine intake is even integrated into the Mediterranean diet, which is widely regarded as a model for healthy eating (6,12). In conditions such as MASLD/NAFLD, alcohol type, dosage, and comorbidities should be considered for personalized clinical evaluation.

While TMREL and NDE offer statistically sophisticated tools for quantifying alcohol-related harm, their clinical and public-health utility is undermined by key limitations as illustrated in Figure 1 (4).

These metrics fail to account for liver-specific outcomes—despite the liver being the primary organ affected by alcohol toxicity—and present age-stratified risk thresholds that complicate public messaging. The abstraction inherent in these models obscures actionable guidance, particularly in hepatology, where alcohol-induced injury is cumulative, organ-specific, and often irreversible (1,4,7,8,11).

Without integration of beverage type, comorbidities

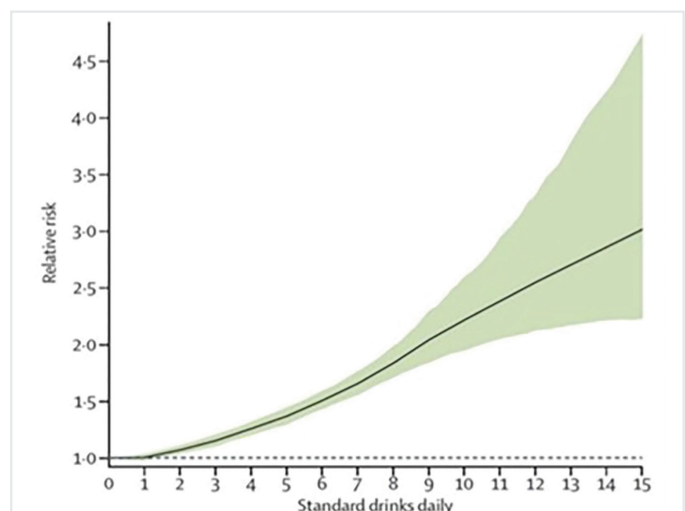


Figure 1. Weighted relative risk of alcohol consumption across all attributable causes by daily standard drink intake, using age-standardized weights determined by the DALY rate in 2016 for both sexes. The dotted line represents the reference point for a relative risk of 1

DALY: Disability-adjusted life-year, Source: Adapted from GBD 2022 (4); created by authors

(e.g., MASLD), and consumption patterns, such models risk oversimplification of complex pathophysiological processes.

In addition, such models rarely consider non-linear hepatotoxic thresholds, where small increases in intake may disproportionately escalate hepatic injury, especially in sensitized populations (1,4,7).

In clinical hepatology, it is frequently observed that individuals reporting what would typically be classified as “moderate” alcohol consumption nonetheless present with elevated transaminase levels, hepatic steatosis on imaging, or even clinical features consistent with alcohol-related liver injury (3,4).

Notably, such presentations often occur at intake levels considered low-risk by population-based models, underscoring a significant disconnect between epidemiological projections and real-world hepatic outcomes. The GBD 2016 analysis (2) visually demonstrates a progressive increase in relative risk for all-cause disease burden with each additional standard drink consumed per day—even within low consumption ranges. These data reinforce the notion that minimal alcohol intake may still pose measurable health risks, particularly to the liver, and call into question the validity of universal “safe” thresholds. This highlights the need for organ-specific, clinically grounded risk-stratification models that better reflect individual vulnerability and hepatotoxic potential (1,4,7,8).

Finally, alcohol use, apart from the liver or cardiovascular system, carries the risk of addiction. Even in moderate use, alcohol can cause dependence in both experimental animals (13) and humans (14), and alcohol addiction remains a significant public-health concern (15,16). The impact of alcohol addiction should be systematically considered when evaluating the full spectrum of alcohol-related health risks. Furthermore, liver and cardiovascular complications progress more rapidly and severely in individuals with alcohol dependence (17,18).

Conclusion

In conclusion, inconsistencies in GBD alcohol studies compromise both the credibility of public messaging and its clinical applicability (1-4). Future guidelines should be anchored in high-certainty evidence—such as RCTs and meta-analyses—and must include hepatology-specific indicators, differentiation by beverage type, and pragmatic risk communication strategies (9,10). Current clinical guidelines such as those jointly issued by EASL, EASD, and EASO, emphasize individual variability in alcohol type, metabolic comorbidities, and drinking patterns; therefore, future GBD updates should align with this multidimensional approach (1,11). Ultimately, risk-stratification models must be organ-specific, evidence-based, and clinically interpretable—reflecting the true complexity of alcohol’s

hepatotoxic potential in populations increasingly affected by MASLD (1,4,7,8,11).

Footnotes

Authorship Contributions

Concept: A.B., Design: A.B., Analysis or Interpretation: A.B., M.U., Literature Search: A.B., M.U., T.U., Writing: A.B., M.U., T.U.

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. Musto JA, Brown R, Lucey MR. Is there a safe limit for consumption of alcohol? *J Hepatol.* 2025;82:535-40.
2. GBD 2016 Alcohol Collaborators. Alcohol use and burden for 195 countries and territories, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet.* 2018;392:1015-35. Erratum in: *Lancet.* 2018;392:1116. Erratum in: *Lancet.* 2019;393:e44.
3. Burton R, Sheron N. No level of alcohol consumption improves health. *Lancet.* 2018;392:987-8.
4. GBD 2020 Alcohol Collaborators. Population-level risks of alcohol consumption by amount, geography, age, sex, and year: a systematic analysis for the Global Burden of Disease Study 2020. *Lancet.* 2022;400:185-235. Erratum in: *Lancet.* 2022;400:358.
5. Radonjić S, Maraš V, Raičević J, Košmerl T. Wine or beer? Comparison, changes and improvement of polyphenolic compounds during technological phases. *Molecules.* 2020;25:4960.
6. Estruch R, Ros E, Salas-Salvadó J, Covas MI, Corella D, Arós F, et al.; PREDIMED Study Investigators. Primary prevention of cardiovascular disease with a mediterranean diet supplemented with extra-virgin olive oil or nuts. *N Engl J Med.* 2018;378:e34.
7. Asrani SK, Devarbhavi H, Eaton J, Kamath PS. Burden of liver diseases in the world. *J Hepatol.* 2019;70:151-71.
8. Narro GEC, Díaz LA, Ortega EK, Garín MFB, Reyes EC, Delfin PSM, et al. Alcohol-related liver disease: a global perspective. *Ann Hepatol.* 2024;29:101499.
9. Ronsley PE, Brien SE, Turner BJ, Mukamal KJ, Ghali WA. Association of alcohol consumption with selected cardiovascular disease outcomes: a systematic review and meta-analysis. *BMJ.* 2011;342:d671.
10. Xi B, Veeranki SP, Zhao M, Ma C, Yan Y, Mi J. Relationship of alcohol consumption to all-cause, cardiovascular, and cancer-related mortality in U.S. Adults. *J Am Coll Cardiol.* 2017;70:913-22. Erratum in: *J Am Coll Cardiol.* 2017;70:1542.
11. Wu XN, Xue F, Zhang N, Zhang W, Hou JJ, Lv Y, et al. Global burden of liver cirrhosis and other chronic liver diseases caused by specific etiologies from 1990 to 2019. *BMC Public Health.* 2024;24:363.
12. Renaud S, de Lorgeril M. Wine, alcohol, platelets, and the French paradox for coronary heart disease. *Lancet.* 1992;339:1523-6.

13. Uzbay IT, Kayaalp SO. A modified liquid diet of chronic ethanol administration: validation by ethanol withdrawal syndrome in rats. *Pharmacol Res.* 1995;31:37-42.
14. Becker HC. Alcohol dependence, withdrawal, and relapse. *Alcohol Res Health.* 2008;31:348-61.
15. Mohapatra S, Patra J, Popova S, Duhig A, Rehm J. Social cost of heavy drinking and alcohol dependence in high-income countries. *Int J Public Health.* 2010;55:149-57.
16. Caetano R, Cunradi C. Alcohol dependence: a public health perspective. *Addiction.* 2002;97:633-45.
17. Lucey MR, Weinrieb RM. Alcohol and substance abuse. *Semin Liver Dis.* 2009;29:66-73.
18. Quintana DS, McGregor IS, Guastella AJ, Malhi GS, Kemp AH. A meta-analysis on the impact of alcohol dependence on short-term resting-state heart rate variability: implications for cardiovascular risk. *Alcohol Clin Exp Res.* 2013;37(Suppl 1):E23-9.



Comparative Efficacy of Diazepam Monotherapy Versus Diazepam-haloperidol Combination in the Treatment of Stimulant-induced Agitation: A Double-blind Clinical Trial

Uyarıcı Kaynaklı Ajitasyonun Tedavisinde Diazepam Monoterapisi ile Diazepam-haloperidol Kombinasyonunun Karşılaştırmalı Etkinliği: Çift Kör Klinik Çalışma

Faraz ZANDIYEH¹, Maral RAMEZANI², Shahin SHADNIA³, Babak MOSTAFAZADEH³, Mitra RAHIMI³, Nava SOLEIMANI¹, Peyman ERFANTALABEVINI³, Eka KURDADZE⁴, Jemima D. Desir DOUGE⁵

¹Shahid Beheshti University of Medical Sciences Faculty of Medicine, Department of Clinical Toxicology, Tehran, Iran

²Arak University of Medical Sciences Faculty of Medicine, Department of Pharmacology, Arak, Iran

³Toxicological Research Center, Excellence Center of Clinical Toxicology, Lohman Hakim Hospital, Shahid Beheshti University of Medical Sciences, Department of Clinical Toxicology, Tehran, Iran

⁴Tbilisi State Medical University Faculty of Medicine, Department of Toxicology, Tbilisi, Georgia

⁵Tampa General Hospital, Florida Poison Information Center, Tampa, Florida, USA

ABSTRACT

Objective: This study aimed to compare the efficacy of diazepam monotherapy versus diazepam combined with haloperidol in managing stimulant-induced agitation.

Methods: A double-blind clinical trial was conducted on 58 patients presenting with agitation due to stimulant intoxication. Agitation severity was assessed using the Richmond Agitation-Sedation Scale at baseline and at 15, 30, 45, and 60 minutes after treatment.

Results: Both treatment regimens significantly reduced agitation within 15 minutes. No statistically significant differences were observed between groups at any time point. Baseline-adjusted analyses confirmed comparable treatment effects.

Conclusion: Diazepam alone and in combination with haloperidol showed similar efficacy in controlling stimulant-

ÖZ

Amaç: Bu çalışmada, uyarıcı madde kullanımına bağlı ajitasyonun yönetiminde diazepam monoterapisi ile diazepam ve haloperidol kombinasyonunun etkinliğinin karşılaştırılması amaçlandı.

Yöntemler: Uyarıcı madde intoksikasyonuna bağlı ajitasyon ile başvuran 58 hastanın dahil edildiği çift kör bir klinik çalışma yürütüldü. Ajitasyon şiddeti, başlangıçta ve tedavi sonrası 15, 30, 45 ve 60. dakikalarda Richmond Ajitasyon-Sedasyon Ölçeği kullanılarak değerlendirildi.

Bulgular: Her iki tedavi rejiminin de 15 dakika içerisinde ajitasyon şiddetini anlamlı düzeyde azalttığı görüldü. Hiçbir zaman noktasında gruplar arasında istatistiksel olarak anlamlı bir fark saptanmadı. Başlangıç değerlerine göre düzeltilmiş analizler de tedavi etkilerinin gruplar arasında benzer olduğunu doğruladı.

Address for Correspondence: Assoc. Prof., Peyman Erfantalabevini, Toxicological Research Center, Excellence Center of Clinical Toxicology, Lohman Hakim Hospital, Shahid Beheshti University of Medical Sciences, Department of Clinical Toxicology, Tehran, Iran

E-mail: peyman1346erfan@gmail.com

ORCID IDs of the authors: F.Z.: 0000-0001-9790-525X, M.R.: 0000-0001-7860-9180, S.S.: 0000-0002-9401-0781, B.M.: 0000-0003-4872-9610, M.R.: 0000-0002-8785-8207, N.S.: 0000-0002-2884-5326, P.E.: 0000-0002-0995-4973, E.K.: 0009-0004-7418-8656, J.D.D.: 0009-0002-1472-933X

Cite this article as: Zandiyeh F, Ramezani M, Shadnia S, Mostafazadeh B, Rahimi M, Soleimani N, et al. Comparative efficacy of diazepam monotherapy versus diazepam-haloperidol combination in the treatment of stimulant-induced agitation: a double-blind clinical trial. Bezmialem Science. 2026;14(4):326-332



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

Received: 08.11.2025

Accepted: 11.03.2026

Published date: 26.10.2026

ABSTRACT

induced agitation. Adverse effects were not systematically assessed; therefore, no conclusion regarding the relative safety of the two regimens can be drawn.

Keywords: Psychomotor agitation, central nervous system stimulants, poisoning, emergency medical services, diazepam, haloperidol

ÖZ

Sonuç: Diazepam monoterapisi ile diazepam ve haloperidol kombinasyonunun uyarıcı madde kullanımına bağlı ajitasyonun kontrolünde benzer etkinlik gösterdiği saptandı. Advers etkiler sistematik olarak değerlendirilmediğinden, iki rejimin göreceli güvenliliği konusunda bir sonuç çıkarılamaz.

Anahtar Kelimeler: Psikomotor ajitasyon, merkezi sinir sistemi uyarıcıları, zehirlenme, acil sağlık hizmetleri, diazepam, haloperidol

Introduction

Agitation is linked to various psychiatric conditions, such as schizophrenia, bipolar disorder, personality disorders (primarily antisocial and borderline), generalized anxiety disorder, panic disorder, major depression, as well as substance use or intoxication (1). Stimulants, substances known for their high potential for addiction, are becoming more common globally (2,3). These substances work by increasing the levels of central monoamines, which are neurotransmitters found in the midbrain and brainstem (2,3). These neurotransmitters are crucial for learning, thinking, problem-solving, and the brain's reward system (2,3). People who have recently used stimulants might show symptoms like restlessness, irritability, euphoria, increased sexual activity, paranoia, delusions, disorganized thinking, or even violent behavior (4).

Emergency room doctors face significant challenges when treating people who use methamphetamine and synthetic cathinones, mainly because of the increased agitation and paranoia these substances cause (5), coupled with the prolonged half-lives of these substances. Methamphetamine can lead to psychotic symptoms (6). These psychotic effects can cause significant agitation and may necessitate immediate medical intervention (5,6).

The standard treatment for this condition typically includes an intramuscular antipsychotic, such as haloperidol (5 mg) or olanzapine (10 mg), a benzodiazepine like lorazepam (2 mg) or midazolam (5 mg), or a combination of both (7). Research indicates that agitation resulting from stimulant use is best managed with high-dose benzodiazepines rather than neuroleptics (4,8). While neuroleptics like haloperidol or olanzapine are typically first-line treatments for agitation and psychosis, their use should be approached cautiously in these patients (4,8-10).

Current trials lack sufficient evidence to confirm or deny the effectiveness of benzodiazepines alone or in combination with other medications for urgent tranquilization or sedation (11). Due to the contradictions in the literature, this study explored and compared two approaches for treating agitated patients poisoned by stimulants: administering diazepam alone versus administering diazepam in combination with haloperidol.

Methods**Study Design and Setting**

A double-blind clinical trial was carried out at the Emergency Department of Loghman Hakim Hospital in Tehran, Iran, from April to July 2025. The study protocol received approval from the Ethics Committee of Shahid Beheshti University of Medical Sciences (approval no: IR.SBMU.RETECH.REC.1402.451, date: 12.11.2023) and was registered with the Iranian Registry of Clinical Trials (IRCT20210720051946N6). Written informed consent was obtained from the patients' legal guardians.

Participants

Eligible participants were adults aged over 18 years who presented with agitation caused by stimulant intoxication, as diagnosed based on DSM-5 criteria (12). Stimulant intoxication was primarily related to methamphetamine use, as confirmed by patient history and clinical presentation. Individuals were excluded if they declined participation or had a history of neurological conditions such as seizures, dementia, or Parkinson's disease; moderate-to-severe head trauma; glaucoma; psychosis unrelated to stimulant use; substance dependence other than stimulants; reduced consciousness; or cardiac issues, including long QT syndrome. Alcohol consumption and previous sedative-hypnotic medication were noted because they may have an impact on the degree of agitation and the pharmacological reaction to benzodiazepines. To ensure the groups had similar starting points, these variables were examined.

Randomization and Intervention

Using a dice-based allocation method, sixty patients were recruited and randomly assigned to two groups (the first group was represented by black dice, and the second group by white dice). Group 1 received diazepam 10 mg (maximum 1 mg/kg). Group 2 received diazepam 10 mg plus haloperidol 5 mg (maximum 30 mg/day). The allocation process was blinded, and outcome assessors were not involved in randomization. Double-blinding in this study meant that both patients and outcome assessors were blinded. An independent clinician who was not involved in agitation assessment or data analysis prepared

and administered the medications. Throughout the study, outcome assessors were not informed of group allocation.

Measures of Outcomes

The Richmond Agitation-Sedation Scale (RASS) was used to assess agitation at baseline and 15, 30, 45, and 60 minutes after treatment began. A validated instrument that is frequently utilized in critical care and emergency situations is the RASS. It has demonstrated excellent inter-rater reliability (intraclass correlation coefficient >0.9) and strong validity for assessing agitation and sedation levels (13). It is also explained in Table 1. Agitation was defined as a RASS score of +2 or higher, and the proportion of patients who remained agitated (RASS ≥+2) at 15 minutes was used as the binary outcome for the risk-based measures described below.

Statistical Analysis

With an 80% power, a 95% confidence level, and an effect size of 0.8, the sample size was calculated using G*Power (version 3.1.9.2), yielding 26 participants per group. Sixty patients were recruited to account for possible dropouts.

The SPSS version 26 was used for data analysis. To determine normality, the Kolmogorov-Smirnov test was used. The chi-square test was used to analyze categorical variables (Fisher’s exact test when expected cell counts were <5), while independent t-tests were used for continuous data, with the Mann-Whitney U test as a sensitivity analysis because RASS is an ordinal scale. Eta squared (η^2) was used to express effect sizes. Additional measures, calculated for the proportion of patients who remained agitated at 15 minutes with the diazepam + haloperidol group as the reference, included the absolute risk reduction (ARR), relative risk (RR), RR reduction (RRR), and number needed to treat (NNT), each with 95% confidence intervals (CIs). A p-value of ≤0.05 was considered statistically significant.

Because the groups started with significantly different levels of agitation, we used an analysis of covariance (ANCOVA) to compare the RASS scores after treatment. This analysis was adjusted for the initial RASS scores. We checked the assumption that the regression slopes were the same, and we found no issues. Adjusted mean values and effect sizes were reported where appropriate.

Results

Based on inclusion and exclusion criteria, 60 patients were divided into two groups. However, two patients were excluded from the study after it was found that they were allergic to diazepam prior to starting treatment. Finally, the study was conducted with 58 patients, with 31 assigned to the diazepam group and 27 to the diazepam plus haloperidol group. One patient in each group had no RASS assessment at 45 and 60 minutes; therefore, analyses at these time points included 30 patients in the diazepam group and 26 in the diazepam plus haloperidol group.

The average ages were 30.5±12.9 years and 27.9±8.9 years, respectively, with no significant differences in demographic characteristics between the groups (p>0.05) (Table 2). A history of alcohol consumption was reported in 3 patients (9.68%) from the diazepam group and 6 patients (22.22%) from the diazepam + haloperidol group (p=0.279, Fisher’s exact test) (Table 2). The history of using sedative-hypnotic drugs was 13 (41.94%) in the diazepam group and 11 (40.74%) in the diazepam combined with haloperidol group (Table 2).

According to Table 1, a patient scoring +2 or higher is classified as agitated, scores from 0 to below +2 are considered normal, and negative scores indicate sedation. The comparison of mean scores between the two groups at various time points is presented in Table 3. According to the RASS, both treatment regimens successfully decreased agitation within 15 minutes. No significant differences were noted between the two groups at 15, 30, 45, or 60-minutes following treatment. The Mann-Whitney U test gave consistent results (p=0.294, 0.264, 0.624, and 0.571 at 15, 30, 45, and 60 minutes, respectively). Mean RASS scores were numerically lower in the diazepam group at all time points, but these differences were not statistically significant. In an exploratory analysis, deep sedation (RASS ≤-3) at 60 minutes was observed in 4 of 30 patients (13.3%) in the diazepam group and 1 of 26 patients (3.8%) in the combination group (Fisher’s exact p=0.358). An ANCOVA was used to compare post-treatment RASS scores while controlling for baseline values because baseline agitation scores varied significantly between groups. There were no statistically significant differences between the diazepam and diazepam plus haloperidol groups at any follow-up time point after adjusting for baseline agitation severity (Table 4).

Table 1. Richmond Agitation-Sedation Scale (RASS) score	
RASS score	Description
+4	Combative, violent, danger to staff
+3	Pulls or removes tube(s) or catheters; aggressive
+2	Frequent non-purposeful movement, fights ventilator
+1	Anxious, apprehensive, but not aggressive
0	Alert and calm
-1	Awakens to voice (eye opening/contact) >10 seconds
-2	Light sedation; briefly awakens to voice (eye opening/contact) <10 seconds
-3	Moderate sedation; movement or eye opening. No eye contact
-4	Deep sedation; no response to voice, but movement or eye opening to physical stimulation
-5	Unarousable; no response to voice or physical stimulation

At 15 minutes, 10 of 31 patients (32.3%) in the diazepam group and 11 of 27 patients (40.7%) in the diazepam + haloperidol group were still agitated (RASS $\geq +2$) (Table 5; Fisher's exact $p=0.588$). With the combination group as the reference, the ARR was 8.5% (95% CI: -16.3% to +33.3%), the RR was 0.79 (95% CI: 0.40-1.57), and the RRR was 20.8% (95% CI: -56.8% to +60.0%). The NNT was approximately 12; however, because the CIs included the null value (the 95% CI of the NNT ranged from an NNT for benefit of 3 through infinity to an NNT for harm of 6), no clear advantage could be demonstrated for either treatment approach. The effect size of the group difference in 15-minute RASS scores was small ($\eta^2=0.023$).

Discussion

In the present study, we evaluated the effectiveness of diazepam alone versus the combination of diazepam and haloperidol in treating poisoned patients. The findings indicated that both medication groups effectively

managed agitated patients. Although entirely by chance, the group receiving diazepam combined with haloperidol had a higher average score on the Richmond scale before treatment began. However, when comparing the averages at 15, 30, 45, and 60-minutes post-treatment, no significant differences were observed between the two groups. In both groups, the average score decreased significantly after 15 minutes. Furthermore, mean RASS scores were negative (indicating sedation) in both groups from 30 minutes onward and were numerically lower with diazepam alone, although the difference was not statistically significant. ANCOVA was used with baseline RASS as a covariate to address the baseline imbalance in agitation severity. This adjusted analysis showed no statistically significant difference between the treatments after accounting for baseline severity. Because the study was not designed to demonstrate equivalence or non-inferiority, these findings indicate that no difference was detected rather than proving that the two regimens are equivalent.

Table 2. Demographic data of patients in two groups

		Diazepam group	Diazepam + haloperidol group	p-value
Gender (number (%))	Male	16 (51.61%)	11 (40.74%)	0.408
	Female	15 (48.39%)	16 (59.26%)	
Age (year) (mean ± SD)		30.52±12.9	27.93±8.9	0.383
History of alcohol use [number (%)]	Yes	3 (9.68%)	6 (22.22%)	0.279
	No	28 (90.32%)	21 (77.78%)	
History of taking sedative-hypnotic medication [number (%)]	Yes	13 (41.94%)	11 (40.74%)	0.927
	No	18 (58.06%)	16 (59.26%)	
The p-value for history of alcohol use was calculated with Fisher's exact test SD: Standard deviation				

Table 3. The analysis of average RASS scores across the two groups at different time intervals

	Diazepam group	Diazepam + haloperidol group	p-value
Richmond scale before treatment	2.65 $\pm$ 0.71	3.07 $\pm$ 0.68	0.022*
Richmond scale 15 minute after treatment	0.65 $\pm$ 1.36	1.07 $\pm$ 1.47	0.252
Richmond scale 30 minute after treatment	-0.94 $\pm$ 1.48	-0.48 $\pm$ 1.50	0.253
Richmond scale 45 minute after treatment	-1.13 $\pm$ 1.78	-1.00 $\pm$ 1.36	0.756
Richmond scale 60 minute after treatment	-1.17 $\pm$ 1.86	-1.04 $\pm$ 1.43	0.776
Values are mean $\pm$ SD. At 45 and 60 minutes, n=30 (diazepam) and n=26 (diazepam + haloperidol) *: $P<0.05$, SD: Standard deviation			

Table 4. Baseline-adjusted comparison of RASS scores between treatment groups

Time point	Adjusted mean RASS (diazepam)	Adjusted mean RASS (diazepam + haloperidol)	Mean difference (95% CI)	p-value
15 min	0.86	0.83	0.03 (-0.63 to 0.70)	0.92
30 min	-0.86	-0.57	-0.28 (-1.10 to 0.53)	0.49
45 min	-1.18	-0.94	-0.24 (-1.14 to 0.66)	0.60
60 min	-1.20	-1.01	-0.19 (-1.14 to 0.76)	0.69
Covariate: Baseline RASS score/model: One-way ANCOVA/significance: $p\leq 0.05$. Mean difference = diazepam minus diazepam + haloperidol; n=56 at 45 and 60 minutes RASS: Richmond Agitation-Sedation Scale, ANCOVA: Analysis of covariance, CI: Confidence interval				

Our results are in line with earlier research showing that benzodiazepines by themselves are useful in treating agitation brought on by stimulant intoxication. Because of their anxiolytic, sedative, and muscle-relaxant properties, benzodiazepines are regarded as first-line treatments for agitation brought on by methamphetamine and related stimulant use, which is primarily mediated by excessive central monoaminergic activity (4,7,10). Klein et al. (14) demonstrated that, without raising the risk of side effects, intramuscular midazolam produced more rapid and effective sedation in agitated emergency department patients than haloperidol and other antipsychotics. Furthermore, benzodiazepine monotherapy is at least as effective as antipsychotic treatment for acute agitation and has fewer extrapyramidal and cardiovascular side effects, according to a Cochrane systematic review by Zaman et al. (11).

Despite the fact that antipsychotic combination therapy is frequently used in emergency situations, there is currently insufficient evidence to support its superiority over benzodiazepine monotherapy. Wilson et al. (15) found that concurrent use of benzodiazepines and antipsychotics might raise the risk of oxygen desaturation and respiratory depression, especially in patients who were intoxicated or who also consumed alcohol. Accordingly, after controlling for baseline agitation severity, the current study did not show any further advantages of combining haloperidol with diazepam. Because adverse effects were not assessed in our study, we cannot draw conclusions regarding the relative safety of the two regimens; the safety concerns noted above derive from previous studies and should be evaluated in future trials.

Managing acute agitation in the emergency room, especially among patients who are intoxicated, is a common challenge for emergency physicians from a clinical standpoint. Non-pharmacological methods are frequently

given priority, and many professionals advise verbal de-escalation or seclusion as first tactics for both sober and drunk people (15). The ideal drug should have a quick onset of action, show little variation in pharmacodynamics and pharmacokinetics between patients, and have a low risk of side effects when pharmacological intervention is required. In order to preserve physical examinations, diagnoses, and subsequent treatment choices, it should also provide tranquilization without causing undue sedation (16).

Haloperidol has been used in studies alone and in combination with other medications to control agitation (17). The combination of haloperidol, lorazepam, and diphenhydramine, as well as the combination of haloperidol and lorazepam, rarely required additional agitation medication (17). However, the combination including diphenhydramine was associated with higher rates of oxygen desaturation, hypotension, increased use of physical restraints, and longer hospital stays (17). Research conducted by Klein et al. (14) revealed that intramuscular midazolam resulted in more effective sedation in agitated emergency department patients within 15 minutes compared to haloperidol, ziprasidone, and possibly olanzapine. Additionally, olanzapine was proved to be more effective than haloperidol. No significant differences in adverse events were observed (14).

In comparison to the haloperidol-promethazine group, Mwankemwa (18) observed that the haloperidol-diazepam group exhibited stronger sedative effects, acted more quickly within 15 minutes, and had lower Positive and Negative Syndrome Scale-Excited Component scores at 15 and 30 minutes. Furthermore, the haloperidol-diazepam group contained more drowsy or stuporous patients at 15, 30, and 60 minutes, whereas the haloperidol-promethazine group contained more alert or drowsy patients with fewer cases of stupor. Asadollahi et al. (16) found that postintervention Agitation-Calmness Evaluation Scale

Table 5. Patient count per RASS scale level across various time points

RASS scale	Diazepam + haloperidol group					Diazepam group				
	Before treatment	After 15 min	After 30 min	After 45 min	After 60 min	Before treatment	After 15 min	After 30 min	After 45 min	After 60 min
+4	7	1	0	0	0	4	0	0	0	0
+3	15	3	1	1	1	12	2	1	1	1
+2	5	7	2	1	1	15	8	2	3	3
+1	0	9	5	1	1	0	8	1	1	1
0	0	0	3	4	3	0	3	6	4	5
-1	0	7	7	6	9	0	10	7	4	3
-2	0	0	9	13	10	0	0	13	14	13
-3	0	0	0	0	0	0	0	0	0	0
-4	0	0	0	0	1	0	0	1	3	4

Counts at 45- and 60-minutes sum to 30 (diazepam group) and 26 (diazepam + haloperidol group) because one patient in each group had no assessment at these time points
RASS: Richmond Agitation-Sedation Scale

scores were 4.73 for valproate and 5.45 for haloperidol ($p=0.028$). Haloperidol caused more intense sedation (36.2% vs. 2.5%) and extrapyramidal symptoms (8.7% vs. 0%) than valproate. Intravenous valproate was effective in reducing agitation with a better safety profile in emergency psychiatry (16).

Wilson's research on drug safety for managing agitation revealed that olanzapine, whether used alone or combined with a benzodiazepine, did not lead to higher rates of hypotension compared to haloperidol. However, the combination of olanzapine and benzodiazepines was linked to reduced oxygen saturation levels compared to haloperidol paired with benzodiazepines in patients who consumed alcohol, but not in those without alcohol use (15).

Diazepam alone was used in very few studies. A study contrasting diazepam and haloperidol was highlighted in an intervention review. The findings showed that, following two days of drug-free observation, 20 newly admitted psychotic patients were administered 20-35 mg of haloperidol intravenously, while another 20 patients received 30-40 mg of diazepam intravenously. Post-treatment assessments conducted at 4 and 24 hours using the Brief Psychiatric Rating Scale and The Clinical Global Impressions demonstrated significant improvement in both groups, with no notable differences observed between the two treatments (11). Additionally, this review showed that although benzodiazepines alone might have fewer adverse effects than older antipsychotics, taking them with other drugs may result in preventable side effects (11). According to prior research and the findings of our study, both methods appear to be effective. Although previous studies suggest that combination treatment may carry additional risks, adverse effects were not assessed in the present study, and the difference in the proportion of patients who remained agitated at 15 minutes (32.3% vs. 40.7%) was not statistically significant.

Study Limitations

This study exclusively utilized the RASS. Employing multiple scoring scales could potentially have yielded more robust results. Additionally, complications, particularly extrapyramidal complications, were regrettably not examined in the patients. The study primarily concentrated on efficacy; therefore, no conclusion can be drawn regarding the relative safety of the two regimens. The findings of this study primarily reflect agitation related to methamphetamine intoxication, which may limit generalizability to other stimulant substances. The study was not designed to demonstrate equivalence or non-inferiority, and RASS data at 45 and 60 minutes were missing for one patient in each group.

Conclusion

Diazepam alone and diazepam combined with haloperidol showed similar efficacy in reducing stimulant-induced agitation in emergency situations, with no statistically significant difference between the groups. Because adverse effects were not systematically assessed, no conclusion regarding the relative safety of the two regimens can be drawn. Larger randomized controlled trials are needed to validate these findings and further investigate safety outcomes.

Ethics

Ethics Committee Approval: The study protocol received approval from the Ethics Committee of Shahid Beheshti University of Medical Sciences (approval no: IR.SBMU.RETECH.REC.1402.451, date: 12.11.2023) and was registered with the Iranian Registry of Clinical Trials (IRCT20210720051946N6).

Informed Consent: Written informed consent was obtained from the patients' legal guardians.

Acknowledgments

The authors thank the Toxicological Research Center (TRC) of Loghman Hakim Hospital, affiliated with the Shahid Beheshti University of Medical Sciences, Tehran, Iran, for supporting this study period.

Declaration on the Use of Artificial Intelligence (AI)

In this study, the Claude artificial intelligence language model (developed by Anthropic) was used to translate the original text into English and to revise and improve the quality of the initial translation. All outputs generated by this tool were reviewed and verified by the authors before being incorporated into the final version of the manuscript. The authors assume full responsibility for the final content.

Footnotes

Authorship Contributions

Surgical and Medical Practices: P.E., Concept: F.Z., P.E., Design: P.E., Data Collection or Processing: F.Z., P.E., Analysis or Interpretation: F.Z., P.E., Literature Search: S.S., B.M., M.R., E.K., J.D.D.D., Writing: F.Z., M.R., P.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. Chan EW, Fagioli A, Möller HJ, Grunze H, Llorca PM, Jaffe RL, et al. Assessment and management of agitation in psychiatry: expert consensus. *World J Biol Psychiatry*. 2016;17:86-128.
2. Li X, Su H, Zhong N, Chen T, Du J, Xiao K, et al. Aberrant resting-state cerebellar-cerebral functional connectivity in methamphetamine-dependent individuals after six months abstinence. *Front Psychiatry*. 2020;11:191.

3. Huang J, Zheng Y, Gao D, Hu M, Yuan T. Effects of exercise on depression, anxiety, cognitive control, craving, physical fitness and quality of life in methamphetamine-dependent patients. *Front Psychiatry*. 2020;10:999.
4. Green AS. Stimulant use disorders and related emergencies. *Substance Use and the Acute Psychiatric Patient: Emergency Management*: Springer; 2019. p.51-68.
5. Nguyen J, Lee S, Ankrah D, Knox E. Evaluating the impact of an emergency department protocol that guides management of methamphetamine-induced agitation and psychosis. *Ment Health Clin*. 2022;12:9-14.
6. Jones CM. Patterns and characteristics of methamphetamine use among adults—United States, 2015-2018. *MMWR Morb Mortal Wkly Rep*. 2020;69:317-23.
7. Wodarz N, Krampe-Scheidler A, Christ M, Fleischmann H, Looser W, Schoett K, et al. Evidence-based guidelines for the pharmacological management of acute methamphetamine-related disorders and toxicity. *Pharmacopsychiatry*. 2017;50:87-95.
8. Penders TM. The syndrome of excited delirium following use of “bath salts”. *J Clin Psychiatry*. 2013;74:518.
9. Benzer TI, Nejad SH, Flood JG. Case records of the Massachusetts General Hospital. Case 40-2013. A 36-year-old man with agitation and paranoia. *N Engl J Med*. 2013;369:2536-45.
10. Richards JR, Albertson TE, Derlet RW, Lange RA, Olson KR, Horowitz BZ. Treatment of toxicity from amphetamines, related derivatives, and analogues: a systematic clinical review. *Drug Alcohol Depend*. 2015;150:1-13.
11. Zaman H, Sampson SJ, Beck AL, Sharma T, Clay FJ, Spyridi S, et al. Benzodiazepines for psychosis-induced aggression or agitation. *Cochrane Database Syst Rev*. 2017;12:CD003079.
12. Cooper R. Diagnosing the diagnostic and statistical manual of mental disorders: Routledge; 2018.
13. Sessler CN, Gosnell MS, Grap MJ, Brophy GM, O’Neal PV, Keane KA, et al. The Richmond Agitation-Sedation Scale: validity and reliability in adult intensive care unit patients. *Am J Respir Crit Care Med*. 2002;166:1338-44.
14. Klein LR, Driver BE, Miner JR, Martel ML, Hessel M, Collins JD, et al. Intramuscular midazolam, olanzapine, ziprasidone, or haloperidol for treating acute agitation in the emergency department. *Ann Emerg Med*. 2018;72:374-85.
15. Wilson MP, MacDonald K, Vilke GM, Feifel D. A comparison of the safety of olanzapine and haloperidol in combination with benzodiazepines in emergency department patients with acute agitation. *J Emerg Med*. 2012;43:790-7.
16. Asadollahi S, Heidari K, Hatamabadi H, Vafaei R, Yunesian S, Azadbakht A, et al. Efficacy and safety of valproic acid versus haloperidol in patients with acute agitation: results of a randomized, double-blind, parallel-group trial. *Int Clin Psychopharmacol*. 2015;30:142-50.
17. Jeffers T, Darling B, Edwards C, Vadieli N. Efficacy of combination haloperidol, lorazepam, and diphenhydramine vs. combination haloperidol and lorazepam in the treatment of acute agitation: a multicenter retrospective cohort study. *J Emerg Med*. 2022;62:516-23.
18. Mwankemwa S. Effectiveness of haloperidol-promethazine versus haloperidol-diazepam combination on rapid tranquilization of acutely agitated patients at Mirembe Mental Health Hospital: a randomized control trial: University of Dodoma (Tanzania); 2022.



The Relationship Between Body Image Satisfaction and Depressive Symptoms in Living Liver Donors

Canlı Karaciğer Donörlerinde Ameliyat Sonrası Beden İmajı Memnuniyeti ile Depresyon Belirtileri Arasındaki İlişkinin İncelenmesi

¹İb Tuğba Nur ÖDEN¹, ²İb Tufan GÜMÜŞ², ³İb Şirin DURMUŞ¹, ⁴İb Alper UĞUZ², ⁵İb Murat ZEYTÜNLÜ³

¹Ege University, Organ Transplantation Institute, İzmir, Türkiye

²Ege University Faculty of Medicine, Department of General Surgery, İzmir, Türkiye

³İzmir University of Economics, Department of General Surgery, İzmir, Türkiye

ABSTRACT

Objective: This study aimed to investigate the relationship between body image satisfaction and depressive symptoms in living liver donors.

Methods: This retrospective cross-sectional study was conducted between November 2022, and August 2023. The sample consisted of 92 individuals who had undergone living donor hepatectomy in the past five years at a university hospital, were over 18 years old, had completed at least one-month post-surgery, and voluntarily agreed to participate. Data were collected using a descriptive characteristics form, a body image questionnaire (BIQ-Tr), and a patient health questionnaire-9 (PHQ-9). Statistical analyses included descriptive statistics, Mann-Whitney U test, Kruskal-Wallis, Spearman correlation, and simple linear regression.

Results: The median BIQ-Tr score was 48.00 [interquartile range (IQR): 12.75], and the median PHQ-9 score was 4.00 (IQR: 5.75). There was no statistically significant relationship between BIQ-Tr or PHQ-9 scores and donor characteristics such as gender, marital status, education, employment status, or recipient age and health status ($p>0.05$). A weak negative correlation was found between BIQ-Tr and PHQ-9 scores ($r=-0.279$, $p=0.004$). BIQ-Tr scores negatively and significantly predicted PHQ-9 scores ($\beta=-0.279$, $p=0.007$). Higher body image satisfaction was associated with lower levels of depressive symptoms. Approximately 8% of the variance

ÖZ

Amaç: Bu çalışmada, canlı karaciğer donörlerinde, ameliyat sonrası beden imajı memnuniyeti ile depresyon belirtileri arasındaki ilişkinin incelenmesi amaçlandı.

Yöntemler: Retrospektif kesitsel tipte olan bu araştırma, Kasım 2022 ve Ağustos 2023 tarihleri arasında yürütüldü. Araştırmanın örneklemini, bir üniversite hastanesinde son beş yıl içinde karaciğer nakli için canlı donör olan, 18 yaş üzeri, canlı donör hepatektomi ameliyatından en az bir ay geçmiş, araştırmaya katılmaya gönüllü olan 92 birey oluşturdu. Veri toplanmasında tanıtıcı bilgiler formu, beden imajı anketi (BİA-Tr) ve hasta sağlık anketi-9 (HAS-9) ölçeği kullanıldı. Verilerin analizinde, tanımlayıcı istatistikler, Mann-Whitney U test, Kruskal-Wallis, Spearman korelasyon, basit doğrusal regresyon testi kullanıldı.

Bulgular: Donörlerin BİA-Tr puan medyanı 48,00 [çeyrekler arası açıklık (ÇAA): 12,75], HAS-9 puan medyanı 4,00 (ÇAA: 5,75) bulundu. Donörlerin cinsiyet, medeni durum, eğitim durumu, çalışma durumu, alıcı yaşı, alıcı sağlık durumu ile BİA-Tr ve HAS-9 puanı arasında istatistiksel olarak anlamlı bir ilişki saptanmadı ($p>0,05$). Donörlerin BİA-Tr puanı ile HAS-9 puanı arasında negatif yönlü zayıf düzeyde istatistiksel olarak anlamlı bir ilişki saptandı ($r=-0,279$, $p=0,004$). BİA-Tr puanı, HAS-9 puanını istatistiksel olarak negatif olarak anlamlı biçimde etkiledi ($\beta=-0,279$, $p=0,007$). Vücut imajından duyulan memnuniyetin yüksek olması, depresif belirtilerin daha düşük

Address for Correspondence: Tuğba Nur Öden PhD, Ege University, Organ Transplantation Institute, İzmir, Türkiye

E-mail: tugba.nur.ozturk@ege.edu.tr

ORCID IDs of the authors: T.N.Ö.: 0000-0002-3460-9725, T.G.: 0000-0003-1861-8282, Ş.D.: 0009-0009-3987-4403, A.U.: 0000-0001-8127-5644, M.Z.: 0000-0001-6911-8953

Cite this article as: Öden TN, Gümüş T, Durmuş Ş, Uğuz A, Zeytüllü M. The relationship between body image satisfaction and depressive symptoms in living liver donors. Bezmialem Science. 2026;14(4):333-340

Received: 20.10.2025

Accepted: 12.03.2026

Published date: 26.10.2026



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

ABSTRACT

in depressive symptoms was accounted for by body image satisfaction ($R^2=0.078$).

Conclusion: Living liver donors had high body image satisfaction and minimal depressive symptoms. Only 6.6% of the donors reported moderate to severe depressive symptoms. Higher body image satisfaction was associated with lower levels of depressive symptoms.

Keywords: Body image, depression, living donors, liver transplantation

Öz

seviyede olmasıyla ilişkilendirildi. Depresif belirtilerdeki varyansın yaklaşık %8'i beden imajından duyulan memnuniyetle açıklanabiliyordu ($R^2=0,078$).

Sonuç: Bu çalışmada karaciğer nakli donörlerinin beden imajı memnuniyeti düzeyleri yüksek, depresyon belirtileri ise minimal düzeyde saptandı. Donörlerin yalnızca %6,6'sında orta şiddetli veya şiddetli düzeyde depresyon belirtisi bulundu. Beden imajı memnuniyeti yüksek olan donörlerin depresyon belirtilerinin düşük olduğu belirlendi.

Anahtar Kelimeler: Beden imajı, depresyon, canlı donörler, karaciğer nakli

Introduction

Liver transplantation is a life-saving treatment for end-stage liver failure. Since organs from deceased donors are not enough to meet the need, liver donation from living donors has become a common alternative (1). Approximately 25% of liver transplants worldwide are performed using living donors, whereas in Türkiye, this rate is around 70% (2,3). According to 2023 data from the Global Observatory on Donation and Transplantation, Türkiye ranks third in the world for living donor liver transplantation (2). At the transplant center where the study was conducted, 56% of liver transplant surgeries are performed using living donors (4).

Body image is a complex concept that integrates an individual's cognitive, perceptual, emotional, and behavioral responses related to their body (5). Body image is a multidimensional construct that encompasses how individuals feel about their bodies, how they perceive their physical appearance, the values and beliefs they hold about their bodies, and the behaviors they engage in to control or manage their body (6). In populations with surgical scars, there is a significant relationship between body image satisfaction and depressive symptoms; however, the magnitude and clinical relevance of this association vary considerably depending on the surgical context. Studies have shown that poorer body image is associated with higher levels of depression across various surgical populations (7-9).

Living liver donors are typically healthy adults who do not require surgical treatment and will not receive any direct medical benefit from the operation. They undergo donor hepatectomy solely on a voluntary basis and must live with a surgical scar in the abdominal area following the procedure (10). This situation may lead to changes in body image after surgery (11). Approximately 42% of living donors may experience changes in body image after surgery, and liver donors have been reported to have lower body image satisfaction compared to kidney donors (12,13).

In recent years, with the increase in living donors in transplantation surgery, patients' body image has gained

greater importance, leading to the adoption of smaller incisions instead of traditional large ones (10,14,15). The literature indicates that donors who undergo minimally invasive surgery report significantly higher levels of body image and satisfaction (16). However, despite their potential benefits, the routine use of minimally invasive techniques in living donor hepatectomy remains limited in many countries (10). In Türkiye, minimally invasive techniques in living donor hepatectomy are applied in a limited number of centers and in selected cases; however, conventional incisions remain the most commonly used approach (17). Although this approach offers many advantages for the surgical team, the abdominal incision may cause mental and psychological stress in donors (16). The surgical scar resulting from living liver donor surgery may affect the donor's post-operative quality of life, overall well-being, and self-confidence (11,18). Depressive symptoms have been reported in approximately 0-34% of patients in the long term after donor surgery (19-22); however, no studies have examined the relationship between depressive symptoms and body image satisfaction.

This study was conducted in Türkiye, where approximately three out of four liver transplants are from living donors, to examine the relationship between postoperative body image satisfaction and depressive symptoms in living liver donors.

Methods**Design and Setting**

This retrospective cross-sectional study was conducted between November 1, 2022, and August 2023 at the organ transplantation department of Ege University Hospital.

Population and Sampling

The study population consisted of 198 individuals who served as living liver donors at the organ transplantation clinic where the study was conducted between October 1, 2017, and 2022 (the past five years before starting the study). The sample size was determined using a priori power analysis conducted with G*Power version 3.1.9.2, based on the effect size reported in the study by Hsu et al.

(23) assuming a 95% confidence level ($\alpha=0.05$), the effect size was set at 0.366, resulting in a minimum required sample size of 91 participants. The study included 92 living liver donors at the specified center over the past five years who underwent donor hepatectomy using the Makuuchi incision (median superior plus right lateral transverse incision), had at least one-month post-hepatectomy, were willing to participate, and had no psychiatric diagnoses.

Data Collection

Data were collected using a data collection form consisting of three sections: a "descriptive characteristics form," "body image questionnaire (BIQ)," and "patient health questionnaire-9 (PHQ-9)".

Descriptive Characteristics Form

The form prepared by the researchers included information on participants' age, gender, occupation, marital status, height, weight, educational level, employment status, and the status of the transplant recipient, among other variables (14,18).

The BIQ

The BIQ was developed by Dunker et al. (24) and has recently been applied in living donor surgeries as well (14,16,25). The BIQ consists of three subscales: body image (5 items, scoring 4 to 20 points), cosmetics (3 items, scoring 3 to 27 points), and self-confidence (2 items, scoring 2 to 20 points). Total scores are calculated for each subscale, with higher scores indicating greater satisfaction with body image. The overall BIQ score ranges from 10 to 67 (19).

Turkish Adaptation of the BIQ (BIQ-Tr)

It is recommended that the sample size for determining the validity and reliability of a measurement tool be five to ten times the number of items in the instrument (26). Since the original BIQ contains 10 items, a sample size of at least 50-100 is recommended to assess its validity and reliability. In this study, the sample size reached 92, which was considered sufficient for validity and reliability evaluation. Details regarding the Turkish adaptation's validity and reliability are as follows:

Language validity

The translation and adaptation process followed the guidelines published by the World Health Organization for instrument translation and validation (27). For the language validity of the instrument, a forward translation was conducted by two healthcare professionals who are native Turkish speakers, proficient in both languages, and familiar with Turkish culture and terminology. The adapted scale was then reviewed by healthcare professionals and a translator. Subsequently, a back-translation was performed by a native English speaker unfamiliar with the original scale. Finally, the two translations were compared to identify any discrepancies.

Content validity

To numerically demonstrate the content validity of the instrument and ensure a reliable evaluation of expert opinions, the Content Validity Index (CVI) was used as the rating criterion (28). For the CVI, evaluations were obtained from three transplant surgeons and three academic nurses. Each expert rated the items on a scale from one to four. Subsequently, the CVI was calculated. A CVI value of at least 0.80 is required to confirm an instrument's content validity (28). In this study, the CVI was found to be 0.93.

Construct validity

Exploratory factor analysis (EFA) was conducted to assess construct validity. Due to correlations among factors, direct oblimin rotation was applied. To determine the suitability of the data, Bartlett's test of sphericity and the Kaiser-Meyer-Olkin (KMO) measure were calculated. KMO values above 0.60 are considered adequate for conducting validity analysis (29). The KMO value in this study was found to be 0.817, indicating the suitability of the instrument for factor analysis. Additionally, Bartlett's test yielded $\chi^2(36)=464.261$, $p=0.000$, demonstrating sufficient inter-item correlations for EFA. The factor analysis revealed that the 9-item BIQ consists of a three-factor structure—body image, cosmetics, and self-confidence—that together explain 76.404% of the total variance. Accordingly, the BIQ-Tr showed valid construct properties. Specifically, the first factor explained 51.587% of the variance, the second 13.661%, and the third 11.157%. Unlike the original BIQ, one item (do you feel less feminine/masculine due to your surgery?) was removed due to cross-loading on multiple factors. The lowest factor loading was 0.707, exceeding the ideal threshold of 0.40, indicating that all items contributed significantly to their respective factors (29). Appendix 1 presents the distribution of items according to factors and their factor loadings.

Reliability

Cronbach's alpha reliability analysis was conducted to assess the instrument's reliability. The reliability coefficients for the subscales were 0.783 for body image, 0.894 for cosmetics, and 0.629 for self-confidence. The overall reliability coefficient of the BIQ-Tr was 0.824, indicating that it is a reliable measurement tool (30). Validity and reliability assessments confirmed that the BIQ-Tr is a valid and reliable measurement tool. The BIQ-Tr consists of three subscales—body image (items 1-4), cosmetics (items 5-7), and self-confidence (items 8-9)—with a total of 9 items. The total possible score on the questionnaire ranges from 10 to 63.

PHQ-9

The PHQ-9 is a measurement tool that assesses nine symptoms of depression based on DSM-IV criteria. The scale was developed by Kroenke et al. (31) and its Turkish validity and reliability were established by Sari et al. (32).

The scale consists of nine items; each scored from zero (not at all) to three (nearly every day). Based on the total score, depression severity is classified as minimal (1-4), mild (5-9), moderate (10-14), moderately severe (15-19), and severe (20-27). The Turkish version of the scale demonstrated a Cronbach's alpha of 0.842 (32). In this study, the scale's Cronbach's alpha was found to be 0.867.

Ethics

Prior to data collection, ethical approval was obtained from the Ege University Medical Research Ethics Committee (approval no: 22-10T/19, date: 06.10.2022) along with written consent from the hospital administration. Prior to participation, all participants were informed about the purpose and procedures of the study, and both written and verbal informed consent was obtained. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Statistical Analysis

Statistical analyses were performed using SPSS version 25.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics, the Shapiro-Wilk test for normality, skewness/kurtosis values, Kruskal-Wallis, Mann-Whitney U test, Spearman correlation, and simple linear regression tests were applied. $P < 0.05$ was considered statistically significant.

Results

Descriptive Characteristics

The mean age of living liver donors at the time of data collection was 36.47 ± 8.66 years. Among the participants, 56.5% were male, 39.2% had bachelor's degree, and 66.3% were married. The average body mass index (BMI) was 24.52 ± 3.37 kg/m², with 54.3% falling within the normal range. Regarding employment status, 57.6% worked full-time, and 41.3% were employed in the private sector. Additionally, 69.6% donated their liver to an adult recipient, and 45.7% voluntarily donated to a parent. The recipients of 20.7% of donors had passed away. Notably, 93.5% of donors stated they would donate again if given the chance (Table 1).

Results Regarding the BIQ-Tr Score

The median BIQ-Tr score among donors was 48.00 [interquartile range (IQR): 12.75]. When examining the

Table 1. Donors' descriptive characteristics (n=92)	
Continuous variables	Mean $\pm$ SD (min-max)
Age (year)	36.47 $\pm$ 8.66 (19-58)
Weight (kg)	71.90 $\pm$ 14.40 (45-118)
Height (m)	1.70 $\pm$ 0.09 (1.55-1.93)
BMI (kg/m ²)	24.52 $\pm$ 3.37 (18.08-33.03)
Pre-tx weight (kg)	70.25 $\pm$ 0.09 (47-104)

Table 1. Continued

Continuous variables	Mean $\pm$ SD (min-max)
Categorical variables	n (%)
Gender	
Female	40 (43.5)
Male	52 (56.5)
Educational status	
Primary or below	23 (25.0)
Middle school	12 (13.0)
High school	21 (22.8)
Bachelor's or above	36 (39.2)
BMI (n=90)	
Underweight	2 (2.2)
Healthy weight	50 (54.3)
Overweight	34 (37.0)
Obesity	4 (4.3)
Marital status	
Single	31 (33.7)
Married	61 (66.3)
Employment status	
Public	14 (15.2)
Private	38 (41.3)
Retired	5 (5.4)
Other*	35 (38.1)
Employment type	
Full-time	53 (57.6)
Part-time	8 (8.7)
Unemployed	31 (33.7)
Relationship with recipient	
Parent	42 (45.7)
Sibling	9 (9.8)
Partner	6 (6.5)
Child	20 (21.7)
Other relatives	15 (16.3)
Recipients' age groups	
Adult	64 (69.6)
Child	28 (30.4)
Recipients' health status	
Alive	73 (79.3)
Deceased	19 (20.7)
Willingness to donate again	
Yes	86 (93.5)
No	6 (6.5)

SD: Standard deviation, Min: Minimum, Max: Maximum, BMI: Body mass index, n: Number, %: Percent, Pre-tx: Pretransplantation, under underweight: <18.5 kg/m², Healthy weight: 18.5-24.9 kg/m², Overweight: 25.0-29.9 kg/m², Obesity: >30 kg, *: Student, homemaker, farmer etc., Other relatives: Aunt/uncle, cousin/nephew etc.

subscales, the median scores were 15.00 (IQR: 2.00) for body image, 15.00 (IQR: 8.00) for cosmetics, and 20.00 (IQR: 4.75) for self-confidence. No statistically significant relationship was found between the median BIQ-Tr score and donors' age, gender, marital status, educational level, BMI, post-transplant weight gain, employment status, kinship with the recipient, recipient's age at transplant, recipient's current health status, or transplant date ($p>0.05$). A statistically significant relationship was found between the donors' status of being a living donor if they had the chance to decide again and the BIQ-Tr score median ($p<0.001$). Individuals who stated that they would want to be a living donor if they had the chance to decide again had a higher BIQ-Tr score median, while those who stated that they would not want to have a lower BIQ-Tr score median (Table 2).

Results Regarding the PHQ-9 Score

The median PHQ-9 score among donors was 4.00 (IQR: 5.75). No statistically significant relationship was found between the median PHQ-9 score and donors' age, gender, marital status, educational level, BMI, post-transplant weight gain, employment status, kinship with the recipient, recipient's age at transplant, or recipient's current health status ($p>0.05$). Moderately severe to severe depressive symptom scores were observed in 6.6% of donors ($n=6$). A statistically significant relationship was found between the donors' decision to become a living donor if they had the chance to decide again and the PHQ-9 median score ($p=0.005$). Individuals who stated that they would not want to become a living donor if they had the chance to decide

again had higher PHQ-9 scores. Additionally, a weak but statistically significant relationship was found between transplant date and the median PHQ-9 score ($p=0.008$), with donors closer to their transplant date exhibiting lower PHQ-9 scores (Table 2).

Results on the Relationship Between BIQ-Tr and PHQ-9 Scores

A weak negative and statistically significant correlation was found between BIQ-Tr and PHQ-9 scores ($r=-0.279$, $p=0.004$). BIQ-Tr scores negatively and significantly predicted PHQ-9 scores ($\beta=-0.279$, $p=0.007$). Higher body image satisfaction was associated with lower levels of depressive symptoms. Approximately 8% of the variance in depressive symptoms was accounted for by body image satisfaction ($R^2=0.078$) (Table 3).

Discussion

Living liver donors in the present study demonstrated high levels of body image satisfaction and minimal depressive symptoms. Moreover, higher body image satisfaction was associated with lower levels of depressive symptoms.

Previous literature has primarily evaluated donor body image satisfaction in relation to incision type or scar appearance. While most studies have found no significant difference in body image scores between conventional and minimally invasive surgery in living liver donors, donors who underwent minimally invasive procedures reported higher levels of cosmetic satisfaction and self-confidence (14,16,23,25). In this study, cosmetic satisfaction

Table 2. Comparison of donors' descriptive characteristics with BIQ-Tr and PHQ-9 scores ($n=92$)

Variables	BIQ-Tr								PHQ-9	
	Body image		Cosmetic		Self-confidence		BIQ-Tr total		r	p
	r	p	r	p	r	p	r	p		
Age	-0.057	0.588	0.132	0.210	-0.058	0.585	0.063	0.553	-0.031	0.769
BMI	-0.155	0.143	-0.092	0.386	-0.800	0.451	-0.090	0.386	0.104	0.326
Transplant date	0.062	0.561	0.028	0.796	0.062	0.562	0.042	0.692	-0.277	0.008*
	Z	p	Z	p	Z	p	Z	p	Z	p
Post-tx weight gain	-0.708	0.479	-0.688	0.491	-0.092	0.927	-0.590	0.555	-0.807	0.420
Gender	-0.227	0.820	-0.257	0.797	-0.060	0.952	-0.363	0.717	-0.657	0.511
Marital status	-0.004	0.997	-1.650	0.099	-0.242	0.809	-1.026	0.305	-0.963	0.336
Recipients' age groups	-1.904	0.057	-1.848	0.065	-0.124	0.901	-1.232	0.218	-0.124	0.902
Recipients' health status	-1.264	0.206	-0.353	0.724	-0.419	0.675	-0.034	0.973	-0.799	0.424
Willingness to donate again	-3.192	0.001*	-3.356	0.001*	-2.722	0.006*	-3.445	0.001*	-2.780	0.005*
	KW	p	KW	p	KW	p	KW	p	KW	p
Educational status	5.011	0.286	1.388	0.846	1.004	0.909	1.273	0.866	2.603	0.626
Employment status	3.126	0.373	2.438	0.487	5.240	0.155	1.849	0.604	1.052	0.789
Employment type	0.134	0.935	0.662	0.718	1.998	0.368	0.825	0.662	1.658	0.437
Relationship with recipient	8.257	0.220	7.960	0.241	4.556	0.602	6.138	0.408	8.605	0.197

BIQ-Tr: Body image questionnaire (Turkish version), PHQ-9: Patient health questionnaire-9, BMI: Body mass index, Post-tx: Post transplantation, r: Spearman correlation test, Z: Mann-Whitney U test, KW: Kruskal-Wallis test, *: $p<0.05$

Table 3. Association between BIQ-Tr scores and PHQ-9 scores

Variables	R ²	F	B	SE	β	p	95% CI	
							Lower	Upper
BIQ-Tr	0.078	7.615	-0.157	0.057	-0.279	0.007	-0.270	-0.044

BIQ-Tr: Body image questionnaire (Turkish version), R²: R square, F: Overall model significance, B: Unstandardized coefficient; SE: Standard error; β: Standardized coefficient, CI: Confidence interval

scores similar to those reported in previous studies for conventional methods were obtained, while body image and self-confidence levels were found to be comparable to results reported for donors undergoing minimally invasive surgery. Overall body image satisfaction scores were also found to be high. These elevated levels of body image satisfaction suggest that donors associate positive meanings with their donation process. Krause et al. (10) reported that living liver donors perceive their surgical process and scars as symbols of pride, and that these scars represent a sign of satisfaction related to their donation experience. On the other hand, the donor process involves an extensive period of medical preparation and counseling (25). During this process, donors are often motivated by the desire to save the recipient's life, and the donor's motivation to donate can influence their body image satisfaction and self-confidence (16,23). Similarly, in this study, donors who stated they would choose to donate again if given the chance demonstrated significantly higher body image satisfaction.

Although living liver donors understand that transplant surgery can potentially save the recipient's life, the negative impact of hepatectomy on their lifestyle may lead to depressive symptoms (16). In this study, donors generally exhibited minimal depressive symptoms, with only 6.6% showing moderately severe to severe signs of depression. Similarly to the present study, in a meta-analysis by Fan et al. (33) including 18,577 living liver donors, the overall prevalence of depression was reported as 6.3%. Furthermore, this study indicates that depression prevalence is higher during the first three months after donation and gradually declines after one year (33). One of the notable findings of the present study is that donors who were closer in time to the transplant surgery exhibited significantly lower levels of depressive symptoms. This finding is consistent with the observations of Hsu et al. (23) and suggests that, in the early post-donation period, donors may experience positive psychological outcomes driven by the motivation of having saved the recipient's life. On the other hand, depressive symptoms in living liver donors have been reported to be influenced by multiple factors, including personal characteristics (female sex, being unmarried, lower educational level), surgery-related factors (serious postoperative complications), recipient-related factors (poor recipient health status or death, negative donor-recipient relationship), and levels of social support (economic difficulties, difficulty finding employment, and inadequate social support) (28). In

this context, the generally minimal level of depressive symptoms observed in our study may be related to participants' personal characteristics (predominantly male and married) and recipient-related factors (the majority of recipients being alive and clinically stable). In addition, the presence of a standardized and comprehensive living donor evaluation protocol at the study center may have facilitated the selection of psychosocially appropriate donors, thereby contributing to these favorable outcomes (34).

Depression has been associated with lower mental quality of life, reduced life satisfaction, and psychological consequences such as post-donation regret (20,35). In the present study, individuals who stated that they would not choose to be a donor if given the opportunity to decide again exhibited significantly higher levels of depressive symptoms. Furthermore, increasing depressive symptoms were associated with decreased body image satisfaction among donors.

Study Limitations

This study has several strengths and limitations. A key strength is that it is one of the few studies conducted in a language for which the BIQ's validity and reliability have been established. On the other hand, the retrospective design of the study is a notable limitation. Prospective studies on body image satisfaction are recommended for future research. Additionally, the absence of a control group and the cross-sectional nature of the study should be considered when interpreting the findings, as these factors limit causal inferences. Moreover, some potential donors may have withdrawn from the donation process during the medical team's evaluation and screening phases due to information provided about scars and possible body image changes. Another limitation is that the study was conducted at a single-center with a limited sample size, which restricts the generalizability of the findings.

Conclusion

This study found that living liver donors generally exhibit high levels of body image satisfaction and minimal depressive symptoms. Donors with higher body image satisfaction demonstrated lower levels of depressive symptoms. Conversely, donors experiencing regret showed lower perceptions of body image, cosmetic satisfaction, and self-confidence, alongside higher depressive symptomatology. Future research is recommended to investigate the impact of decisional conflict prior to donor hepatectomy on post-transplant body image satisfaction and depression.

Ethics

Ethics Committee Approval: Prior to data collection, ethical approval was obtained from the Ege University Medical Research Ethics Committee (approval no: 22-10T/19, date: 06.10.2022).

Informed Consent: Written consent from the hospital administration.

Acknowledgments

We would like to express our sincere gratitude to Veysel Umman and Halit Batuhan Demir for their invaluable contributions in this study.

This study was presented in oral abstract format at the 16th Turkish Hepato-Pancreato-Biliary (HPB) Surgery Congress and the 7th HPB Surgical Nursing Congress (Antalya, November 9-12, 2023).

Footnotes

Authorship Contributions

Surgical and Medical Practices: T.G., A.U., M.Z., Concept: T.N.Ö., T.G., Ş.D., A.U., M.Z., Design: T.N.Ö., T.G., Data Collection or Processing: T.N.Ö., Ş.D., Analysis or Interpretation: T.N.Ö., T.G., Literature Search: T.N.Ö., T.G., Ş.D., Writing: T.N.Ö., T.G.

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

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

Appendix Link: <https://d2v96fxpocvxx.cloudfront.net/d363ec1e-9e5e-4591-a00a-d656bfcabb80/content-images/926334fe-2c0e-49e2-aff9-a50216b57ed7.pdf>

References

- Shah SA, Levy GA, Adcock LD, Gallagher G, Grant DR. Adult-to-adult living donor liver transplantation. *Can J Gastroenterol*. 2006;20:339-43.
- Global Observatory on Donation and Transplantation/ GDOT). International Report On Organ Donation And Transplantation Activities 2023. 2024 Dec. Available from: www.transplant-observatory.org
- Bingol G, Budak Diler S, Sapan I, Dalgic A. Analysis of cadaveric and living donor liver transplantation in Turkey between 2011-2017. *Genel Tip Derg*. 2020;30:119-28.
- Zeytunlu M, Uğuz A, Ünalp Ö, Ergün O, Karasu Z, Günşar F, et al. Results of 1001 liver transplantations in 23 years: Ege University experience. *Turk J Gastroenterol*. 2018;29:664-8.
- Spreckelsen PV, Glashouwer KA, Bennik EC, Wessel I, de Jong PJ. Negative body image: relationships with heightened disgust propensity, disgust sensitivity, and self-directed disgust. *PLoS One*. 2018;13:e0198532.
- Hosseini SA, Padhy RK. Body Image Distortion. StatPearls Publishing; 2023. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK546582/>
- Grant VR, Crerand CE, Jackson JL. Implications of cardiac scarring on body image disturbance among young adults with congenital heart defects. *J Cardiovasc Nurs*. 2024;39:E212-7.
- Thombs BD, Haines JM, Bresnick MG, Magyar-Russell G, Fauerbach JA, Spence RJ. Depression in burn reconstruction patients: symptom prevalence and association with body image dissatisfaction and physical function. *Gen Hosp Psychiatry*. 2007;29:14-20.
- Rhoten BA, Deng J, Dietrich MS, Murphy B, Ridner SH. Body image and depressive symptoms in patients with head and neck cancer: an important relationship. *Support Care Cancer*. 2014;22:3053-60.
- Krause S, Pritlove C, Abbey S, Jung J. The meaning behind the scar: anonymous live liver donors' perceptions of their surgical scars. *Transplantation*. 2021;105:1039-43.
- Zimbrea PC. Body image in transplant recipients and living organ donors. *Curr Opin Organ Transplant*. 2015;20:198-210.
- Trotter JF, Talamantes M, McClure M, Wachs M, Bak T, Trouillot T, et al. Right hepatic lobe donation for living donor liver transplantation: impact on donor quality of life. *Liver Transpl*. 2001;7:485-93.
- DuBay DA, Holtzman S, Adcock L, Abbey SE, Greenwood S, Macleod C, et al. Cosmesis and body image after adult right lobe living liver donation. *Transplantation*. 2010;89:1270-5.
- Suh SW, Lee KW, Lee JM, Choi Y, Yi NJ, Suh KS. Clinical outcomes of and patient satisfaction with different incision methods for donor hepatectomy in living donor liver transplantation. *Liver Transpl*. 2015;21:72-8.
- Dols LF, Ijzermans JN, Wentink N, Tran TC, Zuidema WC, Dooper IM, et al. Long-term follow-up of a randomized trial comparing laparoscopic and mini-incision open live donor nephrectomy. *Am J Transplant*. 2010;10:2481-7.
- Lee JM, Shehta A, Lee KW, Hong SK, Cho JH, Yi NJ, et al. Donor wound satisfaction after living-donor liver transplantation in the era of pure laparoscopic donor hepatectomy. *Surg Endosc*. 2021;35:2265-72.
- Tirnova İ, Alim A, Karataş C, Akbulut A, Demir B, Alper A, et al. Complications of laparoscopic and open donor hepatectomy for living donor liver transplantation: single center experience. *Exp Clin Transplant*. 2024;22:629-35.
- Sinha PK, Mohapatra N, Bharathy KG, Kumar G, Pamecha V. A long-term prospective study of quality of life, abdominal symptoms, and cosmesis of donors after hepatectomy for live-donor liver transplantation. *J Clin Exp Hepatol*. 2021;11:579-85.
- Wang SH, Lin PY, Wang JY, Huang MF, Lin HC, Hsieh CE, et al. Mental health status after living donor hepatectomy. *Medicine (Baltimore)*. 2017;96:e6910.
- Ong JQL, Lim LJH, Ho RCM, Ho CSH. Depression, anxiety, and associated psychological outcomes in living organ transplant donors: a systematic review. *Gen Hosp Psychiatry*. 2021;70:51-75.
- Middleton PF, Duffield M, Lynch SV, Padbury RT, House T, Stanton P, et al. Living donor liver transplantation--adult donor outcomes: a systematic review. *Liver Transpl*. 2006;12:24-30.
- Ng CH, Lim WH, Lim XC, Xiao J, Tan DJH, Syn N, et al. A meta-analysis on the incidence of donor-related depression after liver transplant. *Transpl Int*. 2021;34:2061-70.

23. Hsu YL, Hsieh CE, Lin PY, Lin SL, Lin KH, Weng LC, et al. Postoperative incision scars and cosmetic satisfaction of living liver donors. *Medicine (Baltimore)*. 2021;100:e26187.
24. Dunker MS, Stiggelbout AM, van Hogezaand RA, Ringers J, Griffioen G, Bemelman WA. Cosmesis and body image after laparoscopic-assisted and open ileocolic resection for Crohn's disease. *Surg Endosc*. 1998;12:1334-40.
25. Lind MY, Hop WC, Weimar W, IJzermans JN. Body image after laparoscopic or open donor nephrectomy. *Surg Endosc*. 2004;18:1276-9.
26. Gönül Akkoç M, Aksoy N. Validity and reliability of a Turkish version of the relationship and sexuality scale for kidney transplant recipients. *Akd Med J*. 2021;7:230-8.
27. World Health Organization/WHO. Process of translation and adaptation of instruments. 2017. Available from: http://www.who.int/substance_abuse/research_tools/
28. Davis LL. Instrument review: getting the most from a panel of experts. *Applied Nursing Research*. 1992;5:194-7.
29. Howard MC. A review of exploratory factor analysis decisions and overview of current practices: what we are doing and how can we improve? *Int J Hum Comput Interact*. 2016;32:51-62.
30. Kilic S. Cronbach's alpha reliability coefficient. *Journal of Mood Disorders*. 2016;6:47.
31. Kroenke K, Spitzer RL, Williams JB. The PHQ-9: validity of a brief depression severity measure. *J Gen Intern Med*. 2001;16:606-13.
32. Sari YE, Kokoglu B, Balcioglu H, Bilge U, Çolak E, Ünlüoğlu İ. Turkish reliability of the patient health questionnaire-9. *Biomedical Research-India*. 2016;27:S460-2.
33. Fan Y, Nashan B, Wu X, Zhang H, Shi Y. Prevalence and factors associated with depression and anxiety in living liver donors: a systematic review and meta-analysis. *J Psychosom Res*. 2023;173:111443.
34. Gumus T, Umman V, Sertoz B, Guler E, Uguz A, Sertoz OO, et al. Evaluation of donor exclusions for living donor liver transplantation in a tertiary center. *Transplant Proc*. 2026;58:130-6.
35. Oguten EG, Barlas IS, Akin EB. Mental distress symptoms and life satisfaction among living kidney donors: frequency and association with subjective evaluations. *Transplant Proc*. 2019;51:2232-6.



The Effect of Standardized Patient Simulation on Nursing Students' Insulin Administration Skills

Standardize Hasta Simülasyonunun Hemşirelik Öğrencilerinin İnsülin Uygulama Becerileri Üzerindeki Etkisi

¹Hilal TÜZER¹, ²Bahar VARDAR İNKAYA¹, ³Şenay SARMASOĞLU KILIKÇİER², ⁴Melih ELÇİN³, ⁵Betül ŞAHİN KILINÇ^{4,5}

¹Ankara Yıldırım Beyazıt University Faculty of Health Sciences, Department of Nursing, Ankara, Türkiye

²Mudanya University Faculty of Health Sciences, Department of Nursing, Bursa, Türkiye

³Springfield College, Health Sciences Center, Springfield, Massachusetts, USA

⁴Bağkent University Faculty of Health Sciences, Department of Nursing, Ankara, Türkiye

⁵Ankara Yıldırım Beyazıt University Institute of Health Sciences (PhD), Ankara, Türkiye

ABSTRACT

Objective: Nurses are essential in diabetes care, but both nurses and students often lack knowledge in insulin therapy, which is critical for patient safety given the risks associated with insulin administration. This study aimed to compare the effectiveness of clinical practice alone and clinical practice combined with simulation in improving nursing students' knowledge and skills in managing insulin therapy in real-patient care settings.

Methods: The study was conducted using a quasi-experimental design. In the study, two groups of students, totaling 75 students, were formed and compared: one administered insulin therapy after simulations and interviews, and the other worked directly with real-patients. Both groups were assessed on their insulin therapy skills and knowledge.

Results: Results indicated that the experimental group showed significantly higher knowledge and performance scores than the control group.

Conclusion: The study concluded that clinical practice supported by simulation is more effective than clinical practice

Öz

Amaç: Hemşireler diyabet bakımında kritik bir role sahiptir; ancak hem hemşireler hem de hemşirelik öğrencileri, yüksek riskli bir ilaç olması nedeniyle hasta güvenliği açısından büyük önem taşıyan insülin tedavisi konusunda sıklıkla yetersiz bilgiye sahiptir. Bu çalışma, yalnızca klinik uygulama ile klinik uygulamaya ek olarak simülasyon kullanımının, hemşirelik öğrencilerinin gerçek bakım ortamında insülin tedavisini yönetmeye yönelik bilgi ve becerilerini geliştirmede ne ölçüde etkili olduğunu karşılaştırmayı amaçlamıştır.

Yöntemler: Çalışma, yarı deneysel bir tasarım kullanılarak yürütülmüştür. Çalışmada toplam 75 öğrenciden oluşan iki grup oluşturulmuş ve karşılaştırılmıştır: bir grup, simülasyon uygulamaları ve görüşmeler sonrasında insülin tedavisini uygulamış; diğer grup ise doğrudan gerçek hastalarla çalışmıştır. Her iki grup da insülin tedavisi becerileri ve bilgi düzeyleri açısından değerlendirilmiştir.

Bulgular: Bulgular, deney grubunun bilgi ve performans puanlarının kontrol grubuna kıyasla anlamlı derecede daha yüksek olduğunu göstermiştir.

Address for Correspondence: Betül Şahin Kılınç, Bağkent University Faculty of Health Sciences, Department of Nursing; Ankara Yıldırım Beyazıt University Institute of Health Sciences (PhD), Ankara, Türkiye

E-mail: 215301414@aybu.edu.tr

ORCID IDs of the authors: H.T.: 0000-0002-9929-3688, B.V.İ.: 0000-0002-9151-6795, Ş.S.K.: 0000-0001-9220-5959, M.E.: 0000-0002-1652-906X, B.Ş.K.: 0000-0003-3811-0742

Cite this article as: Tüzer H, Vardar İnkaya B, Sarmasoğlu Kılıkçier Ş, Elçin M, Şahin Kılınç B. The effect of standardized patient simulation on nursing students' insulin administration skills. Bezmialem Science. 2026;14(4):341-348

Received: 01.12.2025
Accepted: 11.05.2026
Published date: 26.10.2026



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

ABSTRACT

alone in enhancing students' insulin therapy management skills, with a direct impact on real patient care.

Keywords: Diabetes mellitus, insulin, nursing education, patient simulation, simulation training, subcutaneous injections

Öz

Sonuç: Çalışma, klinik uygulamanın simülasyon ile desteklenmesinin, öğrencilerin insülin tedavisi yönetimi becerilerini geliştirmede tek başına klinik uygulamaya göre daha etkili olduğunu ortaya koymuştur. Ayrıca bu yaklaşımın gerçek hasta bakımına doğrudan etki sağladığı sonucuna ulaşılmıştır.

Anahtar Kelimeler: Diabetes mellitus, insülin, hemşirelik eğitimi, hasta simülasyonu, simülasyon eğitimi, deri altı enjeksiyonlar

Introduction

The 2025 International Diabetes Federation Diabetes Atlas reports that approximately 589 million adults aged 20 to 79 years are currently living with diabetes globally, accounting for roughly one in every nine adults. Projections indicate that this number will rise to 853 million by the year 2050, reflecting a significant increase in the global disease burden. Alarming, more than 252 million adults are estimated to be living with undiagnosed diabetes, representing over 40% of all cases, which substantially elevates the risk of adverse health outcomes and mortality due to delayed intervention (1). In addition to type 2 diabetes, the global impact of type 1 diabetes (T1D) remains considerable, with an estimated 9.15 million individuals affected across all age groups in 2024, including 1.81 million under the age of 20. Within Europe, Türkiye has the highest adult diabetes prevalence, reported at 16.3%, equating to approximately 9.6 million individuals. This figure positions Türkiye among the top ten countries globally with the highest diabetes burden—a trend projected to continue until at least 2045 (2).

According to projections by Satman et al. (3), over 9.2 million individuals in Türkiye are currently living with diabetes, including approximately 2.6 million older adults. Additionally, an estimated 17 million individuals, of whom 2.5 million are elderly, are affected by some stage of prediabetes. The prevalence of diabetes is anticipated to rise progressively, reaching 17.5% by 2030 and 19.2% by 2045, indicating a consistent upward trend in diabetes incidence across all regions of the country among both men and women in the coming decades (3).

Given the increasing prevalence of diabetes both globally and nationally, effective disease management relies substantially on the clinical competence of healthcare providers—particularly nurses. Both diabetes specialist nurses and general nursing staff hold a central role in improving patient outcomes and overall quality of life, contributing significantly from the early detection of diabetes to the prevention of associated complications (4). Contemporary research highlights that nurse-led care models are not only cost-effective but also associated with improvements in key

clinical indicators, such as glycemic control and decreased hospitalization rates (5). Furthermore, these interventions enhance patients' ability to engage in self-care activities, including medication adherence and lifestyle changes (6). These findings emphasize the critical need for ongoing professional development and institutional support to ensure that nurses are equipped to meet the demands of diabetes prevention and management effectively. Alkubati et al. (7) report that nursing students' inadequate insulin therapy knowledge "negatively affects the care and management of people with diabetes" and therefore recommend that nursing curricula include more detailed instruction on insulin management.

The literature shows that both nurses and nursing students are deficient in diabetes knowledge and management, especially insulin therapy management (8-10). Insulin is one of the most dangerous drugs that threatens the patient's life when administered incorrectly. The Joint Commission on Accreditation of Healthcare Organizations identified insulin as one of the five "high alert" drugs with the highest risk of patient harm due to medication errors (11). The wrong dose of insulin causes symptoms such as hypoglycemia or hyperglycemia in the patient. Such errors during drug administration lead to more severe complications in the future. Therefore, nursing students' knowledge and skills in insulin injection directly affect patient safety.

In traditional education, nursing students receive theoretical education in the classroom, develop their technical skills in the laboratory and learn the art and practice of nursing in the clinical setting. In the clinical setting, students are assigned patients, and the student provides care under the instructor's supervision. Ideally, traditional clinical experiences offer a broad range of learning opportunities. Students learn to think critically, apply their skills, and communicate with the patient, family, and healthcare team. However, the increasing number of undergraduate programs, practices related to patient safety, decreasing length of stay of patients and inequalities in clinical learning experiences have been a challenge for nursing programs (12). At the same time, problems in instructor-student interaction in clinical teaching, inadequate number of

instructors, high number of students, and the lack of clinical teaching environments in which students can acquire the desired skills cause problems in clinical teaching (13,14).

Due to such problems, simulation practice has rapidly increased in nursing education as an innovative and effective educational technique in recent years. Simulation provides students with an active learning experience modeled similarly to real situations. This creates a bridge by filling the gap between theory and practice (15). When the simulation-based training studies on insulin management in the literature were examined, it was observed that students performed this practice with confidence, and it was found that there was a significant increase in their knowledge and performance skills (16-18). The study's findings in which simulation-based training was used to improve nursing care for diabetes showed that perceived glycaemic management knowledge, insulin administration confidence and clinical appropriateness increased after training. These results show that simulation is an effective learning method in insulin management education and helps students gain a better command of diabetes care (19). Simulation-based education creates opportunities for nursing students to develop their knowledge, practice and skills in a safe environment.

Based on these points, this study was planned to compare the performance results of the management of insulin therapy of nursing students in simulation and real-patient care settings.

Study Hypotheses

H1: The students in the experimental group will have significantly higher levels of knowledge regarding insulin therapy compared to those in the control group.

H2: The students in the experimental group will demonstrate significantly higher insulin injection administration skills compared to those in the control group.

Methods

Study Design

The study was conducted using a quasi-experimental design with a control-experimental group post-test design.

Study Sampling and Characteristics

The study was carried out with 3rd semester (experimental group) and 5th semester (control group) students studying in the university's nursing department in Ankara in the 2017-2018 academic year. In the nursing department where the study was conducted, education lasts 4 years and 8 semesters. Seventy-eight of the 3rd-semester students (out of 132 students) and Sixty-four of the 5th-semester students (out of 118 students) who were invited to the study volunteered to participate in the study. From the volunteer students, 40 students were selected from each group using a simple random number table, and the study

was completed with a total of 75 students (control group: 37; experimental group: 38) who fully participated in the study.

Ethical Dimension

In order to evaluate the ethical suitability of the research, the research protocol was submitted for approval by the Clinical Research Ethics Committee of a Ankara Yıldırım Beyazıt University Yenimahalle Education and Research Hospital, and permission was obtained from the relevant ethic committee (protocol no: 2017/38, date: 20.06.2017). Written permission was obtained from the nursing department where the research was conducted. All students and patients who agreed to participate in the study were informed, and the participants signed an informed consent form. Students who wanted to participate in the study but could not participate and who were in the control group were allowed to practice with the standardized patient (SP) after the completion of the study.

Data Collection Forms

The "insulin injection knowledge test" and "insulin injection performance evaluation form" were used for data collection. The researchers developed all data collection tools used in the study, the opinions of the experts for the evaluation of the suitability of measurement tools and five faculty members who are experts in the subject area for the evaluation of content validity were taken and corrections were made according to their suggestions.

Insulin Injection Knowledge Test: This test, prepared to measure the student's knowledge level about insulin injection, consists of 15 multiple-choice questions. Content validity was evaluated by 10 experts in nursing education and clinical assessment using the Davis technique. Items were scored on a four-point relevance scale, with ratings of 3 or 4 accepted as content-valid. The item-level CVI was calculated as the proportion of experts giving these ratings. CVI values ranged from 0.90 to 1.00, supporting the relevance of the items for the target population.

Insulin Injection Performance Evaluation Form: The form prepared to evaluate students' insulin injection performance levels during real-patient care consists of 24 steps, such as controlling blood glucose level, calculating insulin dose, informing the patient, and following up on nutrition. The "not observed", "incomplete/incorrect", and "correct/complete" options were given, and these categories were scored as 0, 1 and 2, respectively. The form's content validity was assessed by the same 10-expert panel using the Davis technique. Item-level CVI values ranged from 0.90 to 1.00, supporting its suitability for evaluating students' clinical examination skills.

Implementation of the Study

The knowledge and performance assessments of the control group students were carried out during the women's health and diseases nursing clinical practice

(5th semester) after two semesters of clinical experience (42 days/336 hours) following the theoretical education in the traditional approach and injection training on the model in the skills laboratory. On the other hand, the experimental group students performed SP interviews and debriefing sessions in addition to the process defined in the traditional method during the internal medicine nursing clinical practice (3rd semester). The experimental group of students worked under the guidance of the instructor for 2 days on subcutaneous injection practice with a low-fidelity simulator before starting clinical practice. In the laboratory study, the practice skills of all students were monitored by the instructor, and individual feedback was given to their performance. One week after the laboratory study, the experimental group students had individual interviews with a SP who simulated the scenario of a patient with diabetes receiving insulin therapy in the simulation laboratory (60 minutes). The scenario includes insulin therapy management, subcutaneous injection knowledge, patient care with diabetes and insulin dose calculation. Simultaneously, the students' practices were recorded on video. Immediately after the practice was completed, a debriefing session (30 minutes) was held with groups of 5 students. In the debriefing session, the video-recorded practices of the students were monitored by both the students and the researchers and were conducted using the Gather-Analyse-Summarise method. It was also possible for the students to evaluate themselves by providing feedback to each other about the deficiencies and performances of the practice.

Students' real-patient insulin administration performance was evaluated in the relevant practice area in the hospital in the last two weeks of clinical practice. Two researchers identified patients receiving insulin therapy in the clinics where the students were present. After obtaining the consent of the patient and the mentor nurse monitoring the practice by the researchers, the students were asked to administer insulin to these patients. Each student's performance was evaluated independently by the same two researchers using the real-patient performance evaluation form.

In the last week of the clinical practice, the control and experimental group students simultaneously participated in the post-test in two separate classrooms with similar characteristics (quiet, comfortable, well-lit). The post-test lasted 20 minutes, and two researchers were present with the students during it (Figure 1).

Statistical Analysis

The IBM SPSS Statistics 21.0 (IBM Corp. Released 2012. IBM SPSS Statistics for Windows, Version 21.0. Armonk, NY: IBM Corp.) program was used for statistical analyses and calculations.

The distribution of knowledge and practice scores in the study was analysed by the Shapiro-Wilk test. All scores

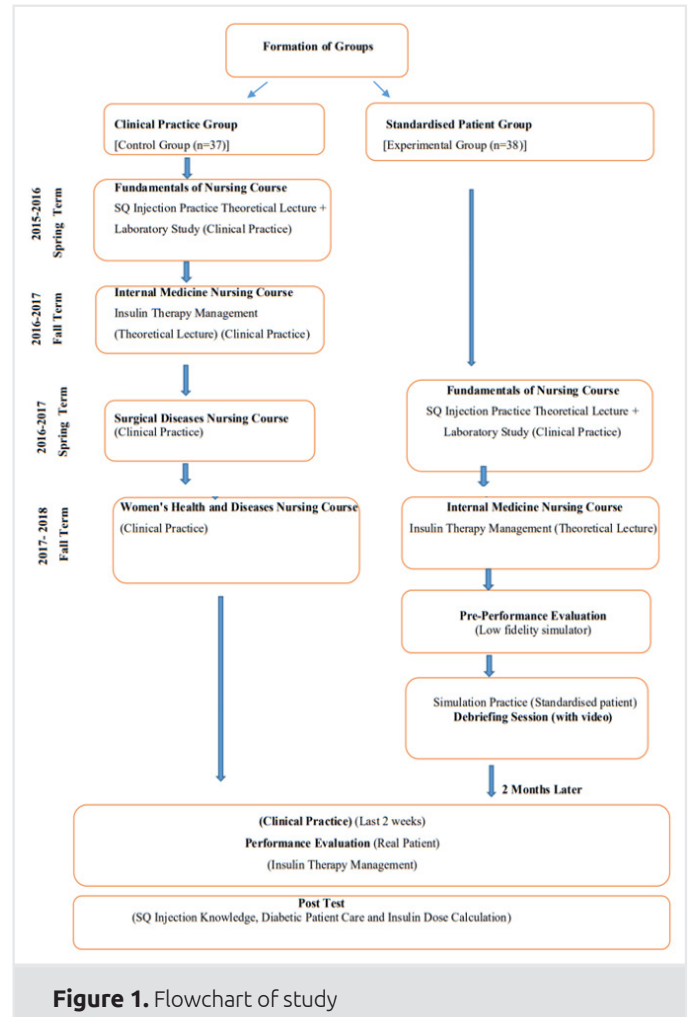


Figure 1. Flowchart of study

were expressed as mean $\pm$ standard deviation. Median (min-max) was given for scores that did not show a normal distribution, and number (%) was given for gender distribution. Knowledge and practice scores were obtained by converting the sum of the scores in the relevant form into a 0-100 scale.

The control and experimental groups were compared with the Mann-Whitney U test regarding post-test knowledge scores and real-patient performance scores. $P < 0.05$ was accepted as statistical significance level. During the evaluations, the researchers did not intervene in student performances in any way. In order to prevent bias, real-patient performance evaluations were collected by two researchers, and the data were coded and transferred to the computer by another researcher without specifying the experimental and control groups.

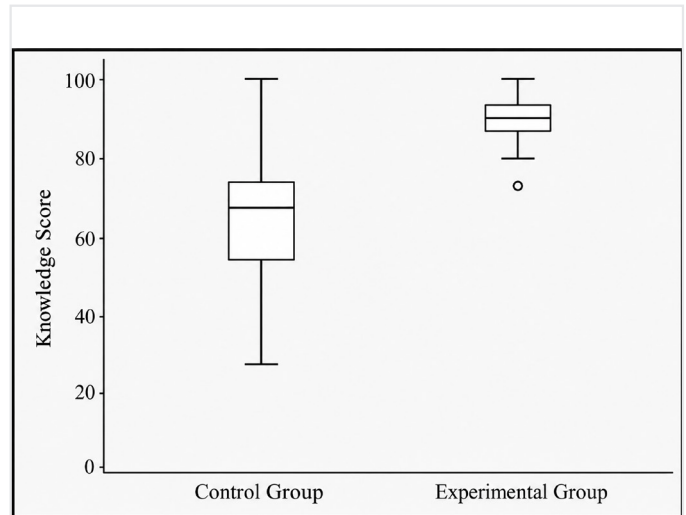
The inter-rater reliability between two observers was assessed by intraclass correlation coefficient (ICC) estimates and their 95% confidence intervals (CIs) based on a single measurement, absolute-agreement, two-way mixed-effects model.

Results

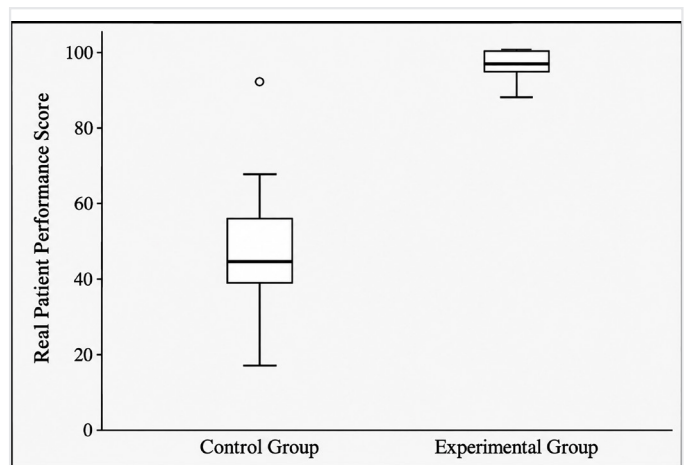
In the study, 86.5% (n=32) of the control group and 84.2% (n=32) of the experimental group were female students. The median age of the control group was 21 (19-25) years, and the experimental group was 20 (19-22) years.

Graphic 1 illustrates the distribution of insulin therapy management knowledge test scores for the control and experimental groups. The results indicate that the experimental group had higher knowledge scores compared to the control group. The median score for the experimental group was notably higher, with a narrower interquartile range, suggesting greater consistency in knowledge acquisition. In contrast, the control group displayed a wider range of scores, with lower median and minimum values. Additionally, the presence of an outlier in the experimental group suggests some variability in the scores, though overall, participants in this group demonstrated superior knowledge compared to those in the control group. These findings suggest that the intervention applied in the experimental group was effective in enhancing the knowledge of the management of the insulin therapy. The median of the insulin therapy management knowledge test of the experimental group was 93.33 (min: 66.67-max: 100.00), which was significantly higher ($p<0.001$) than the control group 66.67 (min: 26.67-max: 100.00) (Table 1).

The median real-patient performance score of the management of the insulin therapy of the experimental group was 95.83 (min: 87.50-max: 100.00). It was significantly higher than the performance score of the control group ($p<0.001$) which was 43.75 (min: 16.67-max: 91.67) (Graphic 2). It was found that statistically significant differences existed between the experimental and control groups across all 24 performance steps ($p<0.05$). These differences were significant at $p<0.001$ for 22 steps, whereas Item 2 (controlling the physician's order) and Item 17 (injecting the drug into the tissue in 10 seconds) showed significance at $p=0.025$ and $p=0.015$, respectively (Supplementary Table S1).



Graphic 1. Distribution of insulin therapy management knowledge test scores in the control and experimental groups



Graphic 2. Control and experimental group insulin therapy management real patient performances

Table 1. Comparison of knowledge and real patient performance scores of the experimental and control groups

	Experimental	Control	p
Knowledge scores			<0.001
Mean $\pm$ SD	90.83 $\pm$ 7.80	65.41 $\pm$ 17.63	
Median (min-max)	93.33 (66.67-100.00)	66.67 (26.67-100.00)	
n	38	37	
Real patient performance scores			<0.001
Mean $\pm$ SD	96.66 $\pm$ 3.23	46.73 $\pm$ 13.31	
Median (min-max)	95.83 (87.50-100.00)	43.75 (16.67-91.67)	
n	38	37	
SD: Standard deviation			

Table 2. Interobserver agreement for real patient performance evaluations based on intraclass correlation coefficients

Type of performance evaluation	Intraclass correlation coefficient	95% confidence interval	Level of agreement	Reference range
Real patient evaluation	0.972	0.950 - 0.984	Excellent	>0.90
Reference range: <0.50, Poor: 0.50-0.75, Moderate: 0.75-0.90, Good: >0.90 excellent				

Two observers independently evaluated participants' insulin administration performance in real-patient care settings. The ICC for their performance scores was 0.972 (95% CI: 0.950-0.984). There results indicated an excellent intra-rater agreement (Table 2).

Discussion

This study was designed to compare the effectiveness of clinical practice following simulation practice with actual clinical practice environments in improving the knowledge and skills of nursing students in insulin therapy management. The results indicated that students with real clinical practice with real-patients scored average on the diabetes mellitus knowledge test. On the other hand, students who had simulation practice before the clinical experience scored significantly higher on the insulin management knowledge test. The findings suggest that simulation-supported clinical practice was associated with higher knowledge and insulin administration performance scores than clinical practice alone.

In the study by Moxley et al. (19), simulation-based didactic education on diabetes care supported significant improvements in students' knowledge, confidence, and clinical practices. This finding aligns with other studies showing that simulation is more effective than traditional teaching methods in fostering clinical skills (20,21). Eom et al. (20) demonstrated that SP-based training improved nursing students'competencies and problem-solving skills. This finding confirms the potential of simulation-based education to bridge the gap between theoretical knowledge and clinical practice (20).

Simulation-based education provides students with a safe opportunity to experience situations they may encounter in real clinical settings, thus helping to bridge the gap between theory and practice. While emphasizing the impact of simulation-based education on the development of clinical skills, this study also highlights that the high-performance scores achieved in real-patient care suggest that simulation-based education is an effective learning tool in laboratory settings and clinical environments. Hayden et al. (12) reported that high-fidelity simulation experiences, accounting for up to half of traditional clinical hours, produced comparable program outcomes and clinical readiness among new graduates. Given that students who participated in simulation demonstrated higher scores in diabetes knowledge and insulin management skills compared to those who only completed clinical internships, the results of this study provide evidence supporting

the value of simulation-based education as an effective teaching strategy, including the use of SPs to enhance students' insulin therapy management skills.

According to our study's findings, students who interacted with SPs obtained significantly higher real-patient insulin therapy management performance scores than students with relatively extended clinical experiences in hospitals. Furthermore, the control group, consisting of students who only participated in clinical internships, exhibited deficiencies in several steps of safe medication administration. These results are consistent with previous studies that highlighted nursing students' inadequate skills in verifying medication rights before administration (22,23), recognizing patients, and performing medication calculations for accurate dosing (24-26).

However, it should also be noted that real clinical experiences should supplement more than simulation-based education. As Fischer-Carlidge et al. (27) notes, simulation exercises may not always fully replace real-patient experiences. Therefore, as in this study, clinical practice supported by simulation exercises emerges as an effective method for maximizing students' knowledge and skills (27).

Furthermore, Whittingham et al.'s (28) study highlighted that a smartphone application developed to strengthen nursing students' clinical skills, such as subcutaneous injections and blood glucose measurement, effectively increased their confidence and performance. These results suggest that technology-enhanced learning supports skill development and enables students to play a more active role in evaluating themselves and their peers. The development of the application through international collaboration underscores the global applicability of technology-based learning across different cultural contexts. Similarly, Skoy et al. (29) conducted a cross-comparison study using two simulators to teach subcutaneous injection techniques. They found that students' confidence and perceived competence increased while anxiety decreased after practicing with both simulators. The only significant difference on performance was observed in students who first practiced with the injection pad and then transitioned to using the simulated arm. Ultimately, students' injection skills and their perceived confidence, competence, and anxiety levels did not vary significantly depending on the type of simulator used (29).

In other studies related to the management of insulin therapy, Ramchandani et al. (30) found that human

patient simulation-based education for parents of newly diagnosed children with T1D provided positive feedback in their diabetes self-management learning process and suggested that this educational model could be extended to other chronic diseases. Faulds et al. (31) also noted that simulation-based education increased accessibility and acceptability among insulin pump users, indicating that this platform could enhance self-management in diabetes care. Similarly, Dubovi et al. (32) found that computer-based simulations had positive effects on diabetes management and understanding of biochemical processes in adolescents with diabetes. Another study testing the applicability and acceptability of simulation in therapeutic patient education (S-TPE) for diabetic adult patients and educators demonstrated that S-TPE is a promising technique for improving diabetes management (33).

In conclusion, this study demonstrates that simulation-based education is an effective method for developing complex skills such as insulin therapy management, as seen in the broader literature. Simulation-based education appears to be an effective teaching strategy that promotes the development of knowledge and skills for a wide range of learners, from nursing students to clinical practitioners and patients. However, it should be acknowledged that simulation is not always the ideal method and must be supported by real clinical practices. This study underscores the need for further research to examine the long-term outcomes of integrating simulation-based technologies and to test their applicability across various disease management scenarios (34,35). Future studies on the long-term impacts of simulation-based education and its practice in different clinical scenarios will further enhance the quality of nursing education and highlight the importance of ongoing research in advancing nursing education.

Study Limitations

Our study was conducted at a single university, and our results cannot be generalized. Since the pre-test scores of the students were not measured, it is impossible to see the initial levels of knowledge and skills regarding insulin management and the progress over time. It is suggested that studies involving a larger sample group and addressing several centers be conducted. In our study, only a single simulation practice has been included since conducting a real-patient performance evaluation was necessary. Students can participate in different simulation practices to compare the differences between clinical practice hours.

Conclusion

The findings indicated that simulation-supported clinical practice was associated with higher knowledge and insulin administration performance scores among nursing students compared with clinical practice alone. This difference was also observed during insulin administration in real-patient care settings. Integrating low-fidelity simulation activities

with SP simulation in the later years of nursing education may provide students with opportunities to update their knowledge and practice their skills in a structured learning environment. These findings reflect improvements in observed student performance rather than patient-level clinical outcomes.

Ethics

Ethics Committee Approval: In order to evaluate the ethical suitability of the research, the research protocol was submitted for approval by the Clinical Research Ethics Committee of a Ankara Yıldırım Beyazıt University Yenimahalle Education and Research Hospital, and permission was obtained from the relevant ethic committee (protocol no: 2017/38, date: 20.06.2017).

Informed Consent: All students and patients who agreed to participate in the study were informed, and the participants signed an informed consent form.

Footnotes

Authorship Contributions

Concept: H.T., M.E., Design: H.T., Ş.S.K., M.E., Data Collection or Processing: H.T., B.V.İ., Ş.S.K., Analysis or Interpretation: B.V.İ., M.E., Literature Search: H.T., B.V.İ., Ş.S.K., B.Ş.K., Writing: H.T., B.V.İ., Ş.S.K., B.Ş.K.

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

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

Supplementary Table Link: <https://d2v96fxpocvxx.cloudfront.net/688d2d00-d207-464d-89b6-73f393f4f50c/content-images/3f53a160-1380-4167-ab98-c3ba25296db2.pdf>

References

1. International Diabetes Federation. The global picture of diabetes. In: IDF Diabetes Atlas. 11th ed. Brussels: International Diabetes Federation; 2025. p.40-71.
2. Ogle GD, Wang F, Haynes A, Gregory GA, King TW, Deng K, et al. Global type 1 diabetes prevalence, incidence, and mortality estimates 2025: results from the International diabetes Federation Atlas, 11th Edition, and the T1D Index Version 3.0. *Diabetes Res Clin Pract.* 2025;225:112277.
3. Satman I, Bayirlioglu S, Okumus F, Erturk N, Yemenici M, Cinemre S, et al.; TURDEP-II Study Group. Estimates and forecasts on the burden of prediabetes and diabetes in adult and elderly population in Türkiye. *Eur J Epidemiol.* 2023;38:313-23.
4. Hu L, Jiang W. Assessing perceptions of nursing knowledge, attitudes, and practices in diabetes management within Chinese healthcare settings. *Front Public Health.* 2024;12:1426339.
5. Zhu Y, Zhang H, Xi Y, Zhu H, Lu Y, Luo X, et al. The implication of diabetes-specialized nurses in aiming for the better treatment and management of patients with diabetes mellitus: a brief narrative review. *Diabetes Ther.* 2024;15:917-27.

6. Alshammari M, Windle R, Bowskill D, Adams G. The role of nurses in diabetes care: a qualitative study. *Open J Nurs.* 2021;11:682-95.
7. Alkubati SA, Albagawi B, Alharbi TA, Alharbi HF, Alrasheeday AM, Llega J, et al. Nursing internship students' knowledge regarding the care and management of people with diabetes: a multicenter cross-sectional study. *Nurse Educ Today.* 2023;129:105902.
8. Yacoub MI, Demeh WM, Darawad MW, Barr JL, Saleh AM, Saleh MY. An assessment of diabetes-related knowledge among registered nurses working in hospitals in Jordan. *Int Nurs Rev.* 2014;61:255-62.
9. Rubin DJ, Moshang J, Jabbour SA. Diabetes knowledge: are resident physicians and nurses adequately prepared to manage diabetes? *Endocr Pract.* 2007;13:17-21.
10. Tawalbeh L, Gharaibeh B. Diabetes knowledge among university nursing students in the north of Jordan. *Pak J Nutr.* 2014;13:728-34.
11. Jagodage H, Sandamali P, Lenora J. Knowledge, attitude and practices on insulin therapy among patients with type 1 diabetes. *Galle Med J.* 2023;28:30-6.
12. Hayden JK, Smiley RA, Alexander M, Kardong-Edgren S, Jeffries PR. The NCSBN National Simulation Study: a longitudinal, randomized, controlled study replacing clinical hours with simulation in prelicensure nursing education. *J Nurs Regul.* 2014;5(2 Suppl):S3-64.
13. Ross JG. Simulation and psychomotor skill acquisition: a review of the literature. *Clin Simul Nurs.* 2012;8:e429-35.
14. Öztürk D, Dinç L. Effect of web-based education on nursing students' urinary catheterization knowledge and skills. *Nurse Educ Today.* 2014;34:802-8.
15. Gaba DM. The future vision of simulation in health care. *Qual Saf Health Care.* 2004;13(Suppl 1):i2-10.
16. Dobbs C, Sweitzer V, Jeffries P. Testing simulation design features using an insulin management simulation in nursing education. *Clin Simul Nurs.* 2006;2:e17-22.
17. Sperl-Hillen J, O'Connor P, Ekstrom H, Rush W, Asche S, Fernandes O, et al. Using simulation technology to teach diabetes care management skills to resident physicians. *J Diabetes Sci Technol.* 2013;7:1243-54.
18. Wilson RD, Bailey M, Boyle ME, Seifert KM, Cortez KY, Baker LJ, et al. Description and preliminary evaluation of a diabetes technology simulation course. *J Diabetes Sci Technol.* 2013;7:1561-6.
19. Moxley E, Maturin LJ, Habtezgi D. A lesson involving nursing management of diabetes care: incorporating simulation in didactic instruction to prepare students for entry-level practice. *Teach Learn Nurs.* 2021;16:10-5.
20. Eom MR, Kim HS, Kim EK, Seong K. [Effects of teaching method using standardized patients on nursing competence in subcutaneous injection, self-directed learning readiness, and problem solving ability]. *J Korean Acad Nurs.* 2010;40:151-60. Korean.
21. Sarmasoglu S, Dinc L, Elcin M. Using standardized patients in nursing education: effects on students' psychomotor skill development. *Nurse Educ.* 2016;41:E1-5.
22. Sulosaari V, Kajander S, Hupli M, Huupponen R, Leino-Kilpi H. Nurse students' medication competence--an integrative review of the associated factors. *Nurse Educ Today.* 2012;32:399-405.
23. Wolf ZR, Hicks R, Serembus JF. Characteristics of medication errors made by students during the administration phase: a descriptive study. *J Prof Nurs.* 2006;22:39-51.
24. Grandell-Niemi H, Hupli M, Leino-Kilpi H, Puukka P. Finnish nurses' and nursing students' pharmacological skills. *J Clin Nurs.* 2005;14:685-94.
25. Zahara-Such RM. Improving medication calculations of nursing students through simulation: an integrative review. *Clin Simul Nurs.* 2013;9:e379-83.
26. Güneş UY, Baran L, Yılmaz DK. Mathematical and drug calculation skills of nursing students in Turkey. *Int J Caring Sci.* 2016;9:220-7.
27. Fischer-Carlidge E, Romanoff S, Thom B, Burrows Walters C. Comparing self-injection teaching strategies for patients with breast cancer and their caregivers: a pilot study. *Clin J Oncol Nurs.* 2016;20:515-21.
28. Whittingham K, Wharrad H, Rutt J, Urstad KH, Strømme T, Mordt P, et al. An evaluation of student nurses' experiences of a smartphone app aimed at developing confidence and competence in performing clinical skills. In: *INTED2020 Proceedings. IATED; 2020. p.3063-4.*
29. Skoy ET, Eukel HN, Frenzel JE. Comparison of low- and higher-fidelity simulation to train and assess pharmacy students' injection technique. *Am J Pharm Educ.* 2013;77:33.
30. Ramchandani N, Maguire LL, Stern K, Quintos JB, Lee M, Sullivan-Bolyai S. PETS-D (parents education through simulation-diabetes): parents' qualitative results. *Patient Educ Couns.* 2016;99:1362-7.
31. Faulds ER, Rayo M, Lewis C, Noble CW, Gifford R, Happ MB, et al. Simulation platform development for diabetes and technology self-management. *J Diabetes Sci Technol.* 2022;16:1451-60.
32. Dubovi I, Levy ST, Levy M, Zuckerman Levin N, Dagan E. Glycemic control in adolescents with type 1 diabetes: are computerized simulations effective learning tools? *Pediatr Diabetes.* 2020;21:328-38.
33. Pennecot C, Luu M, Marchand C, Gagnayre R, Dechannes N, Rudoni S, et al. First use of simulation in therapeutic patient education (S-TPE) in adults with diabetes: a pilot study. *BMJ Open.* 2022;12:e049454.
34. Alaklabi A, Alharbi A, McFeely C, Mullen R, Miller W. Exploring the impact of simulation-based learning (SBL) on the acquisition of psychomotor skills, knowledge, retention, satisfaction and confidence in teaching urinary catheterization in nursing education: systematic review. *Clin Simul Nurs.* 2025;109:101851.
35. Blaak H, Lkoul A, Iziki H, Arechkik A, Sbair Idrissi K, El Hilali S, et al. Effectiveness of simulation with a standardized patient on knowledge acquisition, knowledge retention, and self-efficacy among Moroccan nursing students: a quasi-experimental study. *Healthcare (Basel).* 2025;13:318.



Evaluation of the Prevalence of Mandibular Canal Variations Using Cone-beam Computed Tomography

Mandibular Kanal Varyasyon Prevalansının Konik Işınlı Bilgisayarlı Tomografi ile Değerlendirilmesi

Elifhan ALAGÖZ¹, Nilüfer KARAÇAY¹, Nilüfer GÜRSOY², İrfan SARICA¹

¹Bezmialem Vakıf University Faculty of Dentistry, Department of Oral and Maxillofacial Radiology, İstanbul, Türkiye

²Bahçeşehir University Faculty of Dentistry, Department of Oral and Maxillofacial Radiology, İstanbul, Türkiye

ABSTRACT

Objective: Hidden anatomical variations of the mandibular canal may substantially increase the risk of hemorrhage, neurosensory injury, and surgical complications during implant placement, third molar surgery, and other invasive procedures. Despite their clinical relevance, these variations are frequently overlooked on routine imaging. This study aimed to determine the prevalence and distribution of mandibular canal variations using cone-beam computed tomography (CBCT) and to evaluate their associations with age, sex, edentulism, and laterality.

Methods: In this retrospective cross-sectional study, 794 hemi-mandibular CBCT images from 426 patients (286 females, 140 males; mean age: 43.68±12.34 years) were analyzed. Variations were classified according to the Naitoh system. Associations were assessed using Pearson's chi-square test ($p<0.05$).

Results: Mandibular canal variations were identified in 86.4% of hemi-mandibles. The most prevalent subtype was Type 3 (retromolar canal) (46.2%), followed by Type 2 (dental canal) (28.7%). No significant association was found with sex ($p=0.123$) or side distribution ($p=0.120$). Significant associations were observed with age ($p=0.001$), with Type 3 variations predominating in the 41-50 years group, and with edentulism status ($p=0.001$), particularly in individuals with posterior molar tooth loss.

Conclusion: Mandibular canal variations are highly prevalent and should be regarded as clinically relevant rather than incidental findings. The retromolar canal was the dominant

ÖZ

Amaç: Mandibular kanalın gizli anatomik varyasyonları, implant yerleştirilmesi, üçüncü molar cerrahisi ve diğer invaziv işlemler sırasında hemoraji, nörosensoriyel yaralanma ve cerrahi komplikasyon riskini belirgin şekilde artırabilir. Klinik önemlerine rağmen bu varyasyonlar rutin görüntülemelerde sıklıkla gözden kaçmaktadır. Bu çalışmanın amacı, konik ışınlı bilgisayarlı tomografi (KİBT) ile mandibular kanal varyasyonlarının prevalansını ve dağılımını belirlemek; ayrıca yaş, cinsiyet, dişsizlik durumu ve taraf ile ilişkilerini değerlendirmektir.

Yöntemler: Bu retrospektif kesitsel çalışmada, 426 hastaya ait 794 hemimandibular KİBT görüntüsü (286 kadın, 140 erkek; ortalama yaş: 43,68±12,34 yıl) incelendi. Varyasyonlar Naitoh sınıflamasına göre değerlendirildi. İlişkiler Pearson ki-kare testi ile analiz edildi ($p<0,05$).

Bulgular: Mandibular kanal varyasyonları hemimandibularların %86,4'ünde saptandı. En sık görülen alt tip Tip 3 (retromolar kanal) (%46,2) olup, bunu Tip 2 (dental kanal) (%28,7) izledi. Cinsiyet ($p=0,123$) ve taraf dağılımı ($p=0,120$) ile anlamlı ilişki bulunmadı. Yaş ile anlamlı ilişki saptandı ($p=0,001$); Tip 3 varyasyonlar en sık 41-50 yaş grubunda görüldü. Ayrıca dişsizlik durumu ile de anlamlı ilişki bulundu ($p=0,001$), özellikle posterior molar diş kaybı olan bireylerde daha belirgindi.

Sonuç: Mandibular kanal varyasyonları oldukça yaygın olup tesadüfi bulgular olarak değil, klinik açıdan önemli anatomik oluşumlar olarak değerlendirilmelidir. Retromolar kanal baskın alt tip olarak belirlenirken, yaş ve dişsizlik varyasyon

Address for Correspondence: Asst. Prof., Elifhan Alagöz, Bezmialem Vakıf University Faculty of Dentistry, Department of Oral and Maxillofacial Radiology, İstanbul, Türkiye

E-mail: elifhan.alagoz@bezmialem.edu.tr

ORCID IDs of the authors: E.A.: 0000-0001-9065-5091, N.K.: 0000-0002-3432-1123, N.G.: 0000-0001-5056-7679, İ.S.: 0000-0003-1038-8275

Cite this article as: Alagöz E, Karaçay N, Gürsoy N, Sarica İ. Evaluation of the prevalence of mandibular canal variations using cone-beam computed tomography. Bezmialem Science. 2026;14(4):349-356



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

Received: 26.04.2026

Accepted: 05.06.2026

Published date: 26.10.2026

ABSTRACT

subtype, while age and edentulism significantly influenced variation patterns. Preoperative CBCT assessment is strongly recommended to optimize treatment planning and minimize neurovascular complications.

Keywords: Mandibular canal variations, CBCT, retromolar canal, Naitoh classification

Öz

paternlerini anlamlı şekilde etkilemiştir. Tedavi planlamasını optimize etmek ve nörovasküler komplikasyonları azaltmak için işlem öncesi KIBT değerlendirmesi şiddetle önerilmektedir.

Anahtar Kelimeler: Mandibular kanal varyasyonları, KIBT, retromolar kanal, Naitoh sınıflaması

Introduction

The mandibular canal is defined as a bilateral vascular-neural structure that begins at the mandibular foramen, curves downward and forward beneath the roots of the molar teeth, and terminates at the mental foramen (1). This canal contains the inferior alveolar nerve, inferior alveolar artery, and inferior alveolar vein. It plays a critical role in the functional and anatomical integrity of the mandible (2,3). The location and structure of the mandibular canal are important for endodontic treatment, impacted third molar extraction, dental implant applications, and orthognathic surgery involving the mandible (4-6).

Anatomical variations of the mandibular canal can present a wide range. Variations frequently reported in the literature include bifid mandibular canals, where the canal follows a double path, as well as retromolar canals, forward canals, dental canals, and buccolingual canals (7,8). Accurate identification of the localization of mandibular canal variations not only protects the patient from various surgical complications but also facilitates the clinician's work. These anatomical variations should be examined in detail to reduce the risk of nerve damage and prevent complications (9-12).

In recent years, advancements in dental imaging technologies have made it possible to examine the mandibular canal in greater detail. Conventional radiographic methods provide only two-dimensional imaging of the mandibular canal, making it insufficient to adequately assess depth information or variations (3,13). In contrast, cone-beam computed tomography (CBCT) has become a crucial tool for the detailed evaluation of dental and maxillofacial structures, allowing for three-dimensional and high-resolution imaging of the mandibular canal (10,14,15). The detailed images provided by CBCT enable high-resolution three-dimensional visualization of the mandibular canal and offer significant advantages in detecting canal variations and determining their prevalence (9,11,16).

This study aims to investigate the variations of the mandibular canal using CBCT and evaluate their prevalence. By addressing gaps in the literature and aiming to better understand the clinical significance of mandibular canal

anatomy, this study seeks to provide important insights to guide surgical and dental treatment planning.

Methods**Study Design and Ethical Approval**

This retrospective cross-sectional investigation was approved by the Non-Interventional Ethics Committee of Bezmialem Vakıf University (approval no: E-54022451-050.05.04-108546, decision no: 2023/177, date: 26.05.2023). All study procedures complied with the ethical principles of the Declaration of Helsinki.

Study Population and Sample Selection

This cross-sectional retrospective study utilized CBCT images obtained from patients who attended Bezmialem Vakıf University, Faculty of Dentistry, Department of Oral and Maxillofacial Radiology between September 2022 and December 2022. Images had been acquired for various diagnostic and treatment-planning purposes, including impacted tooth evaluation, implant planning, orthodontic assessment, and orthognathic surgical planning.

All available CBCT scans meeting the inclusion criteria during the study period were screened, aiming to reflect the study population routinely referred for CBCT examination. Inclusion criteria required high-quality CBCT images that fully demonstrated the mandibular canal entrance and course. Exclusion criteria included congenital anomalies or severe jaw deformities, maxillofacial fractures, neoplasms, intraosseous lesions affecting the mandibular canal, and scans degraded by motion or imaging artifacts.

A total of 965 patients' CBCT images were screened. After applying the inclusion and exclusion criteria, CBCT images of 58 patients with unilateral scans and 368 patients with bilateral scans, resulting in a total of 794 CBCT images from 426 patients, were included in the study.

CBCT Acquisition Parameters

All scans were obtained using the Planmeca ProMax 3D Mid Proface unit (Finland) under high-resolution settings (90 kVp, 10 mA, exposure time: 18-36 s; voxel size: 0.2 mm³). Image datasets were reconstructed using Romexis software (version 3.8.3.R) and exported in DICOM format for evaluation.

Image Evaluation and Observer Reliability

All CBCT images were independently evaluated by two dentomaxillofacial radiologists with 3 years (NG) and 4 years (NK) of experience. In cases of disagreement, the images were re-evaluated jointly and a final consensus decision was reached. Observer agreement was assessed using weighted kappa statistics. Inter-observer agreement was calculated based on the initial independent evaluations, and the agreement between each observer's initial assessment and the final consensus evaluation was also analyzed.

Image Reconstruction and Assessment Procedure

Cross-sectional and panoramic images were reconstructed for all hemi-mandibles based on axial sections. Evaluations were performed on cross-sectional slices with a 0.5 mm slice interval. For each hemi-mandibular canal, reconstructed panoramic views and buccolingual cross-sectional slices were created and analyzed using Romexis version 3.8.3.R software.

Panoramic views were reconstructed based on an oblique line drawn through the center of each root on an axial section passing through the cervical regions of the teeth. Perpendicular to this oblique line, 0.5 mm thick buccolingual cross-sectional slices were prepared.

Classification of Mandibular Canal Variations

The prevalence of mandibular canal variations was determined according to the Naitoh classification (16). This classification is categorized as follows (Figure 1):

Type 1 Forward canal

A: Intersecting

B: Non-intersecting

Type 2 Dental canal

C: Branch reaching the first molar

D: Branch reaching the second molar

E: Branch reaching the third molar

Type 3 Retromolar canal

F: Retromolar canal

Type 4 Buccolingual canal

G: Branch diverging from the buccal wall of the main canal

H: Branch diverging from the lingual wall of the main canal.

Outcome Measures

In each hemi-mandible, the prevalence and type of mandibular canal variation were evaluated. The correlations between variation types and age groups, gender, edentulism status, and orientation were also assessed.

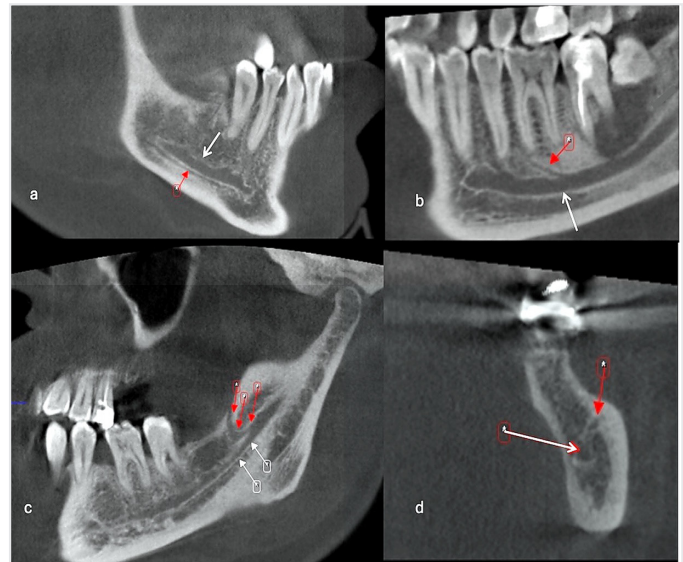


Figure 1. Representative CBCT images demonstrating mandibular canal variations according to the Naitoh classification. (a) Type 1 forward canal (intersecting): accessory canal arising from the main mandibular canal and coursing anteriorly with intersection. (b) Type 2 dental canal: branch extending toward the first molar region. (c) Type 3 retromolar canal: accessory branch extending superiorly toward the retromolar area posterior to the third molar region. (d) Type 4 buccolingual canal (lingual type): branch diverging from the lingual wall of the main mandibular canal. Red arrows indicate the variation canals, whereas white arrows indicate the main mandibular canal

CBCT: Cone-beam computed tomography

Statistical Analysis

All statistical analyses were performed using SPSS version 28.1 (IBM Corp., Chicago, IL, USA). Descriptive variables were summarized as mean, standard deviation, minimum, maximum, frequency, and percentage where appropriate. Associations between categorical variables were tested using Pearson's chi-square test. To account for the potential correlation between the right and left hemi-mandibles within the same individual, an additional analysis was performed using generalized estimating equations (GEE). Weighted kappa coefficients were used to determine observer agreement. A p-value below 0.05 was considered statistically significant.

Results

Study Population Characteristics

A total of 794 hemi-mandibular CBCT images obtained from 426 patients were included in the study. The mean age of the study population was 43.68 ± 12.34 years (range: 10-77 years). Of the evaluated hemi-mandibles, 67.1% (n=533) belonged to female participants and 32.9% (n=261) to male participants. No missing data were present in the final analysis.

Regarding age distribution, the largest proportion of participants was in the 41-50 years age group (25.9%), followed by the 31-40 years group (19.8%). In terms of dentition status, 32.9% of participants had no molar tooth loss, whereas 21.4% had two missing molars. Side distribution showed a slight predominance of the right side (51.5%). Detailed demographic and clinical characteristics are presented in Table 1.

Prevalence and Distribution of Mandibular Canal Variations

Mandibular canal variations were identified in 86.4% of the evaluated hemi-mandibles. The most common variation type was Type 3 (retromolar canal) (46.2%), followed by Type 2 (dental canal) (28.7%). No variation was observed in 13.6% of cases. Detailed distributions of variation types and subtypes are presented in Tables 2 and 3.

Association Between Mandibular Canal Variations and Sex or Side

No statistically significant association was found between sex and mandibular canal variations (χ^2 test, $p=0.123$), indicating a similar distribution between males and females. Likewise, no significant difference was observed

between the right and left sides (χ^2 test, $p=0.120$). When the potential correlation between the right and left hemi-mandibles of the same individual was taken into account using a GEE model, no significant association was observed between side and Naitoh classification ($p=0.364$).

Association Between Mandibular Canal Variations and Age Group

A statistically significant association was identified between age group and mandibular canal variations (χ^2 test, $p=0.001$). Type 3 variations were more frequently observed in the 41-50 years age group. Detailed comparisons are presented in Table 2.

Association Between Mandibular Canal Variations and Edentulism Status

Edentulism status was also significantly associated with mandibular canal variations (χ^2 test, $p=0.001$). Type 3 variation was observed in 19.9% ($n=57$) of individuals with one missing molar and 21.4% ($n=100$) of those with two missing molars. Detailed comparative analyses are presented in Table 2.

Observer Agreement

Observer agreement results are summarized in Table 4.

Discussion

Identification of anatomical variations of the mandibular canal is of vital importance in dental surgery and implantology procedures. In the present study, the prevalence and types of mandibular canal variations were evaluated and the findings were compared with the current literature.

In our study, mandibular canal variations were detected in 86.4% of the 794 hemi-mandibular images examined. This rate is higher than those reported in previous studies. Doğan and Kotanlı (17), reported a prevalence of 66.1% for mandibular canal variations in a Turkish population using CBCT. Similarly, Orhan et al. (18), in their study evaluating root canal morphology of the lower anterior teeth using CBCT, reported a prevalence of 55.6% for mandibular canal variations. Özarslantürk and Avsever (19), on the other hand, found a variation rate of 8.57% in CBCT images of a Turkish population. In a systematic review conducted by Samieirad et al. (20), the overall prevalence of bifid mandibular canal was reported as 18%; however, this rate increased to 25% in studies evaluated using CBCT. The differences observed between the present study and the literature may be attributed to variations in sample size, population characteristics, and imaging techniques used among studies. The relatively high prevalence observed in the present study may also be related to methodological factors. The use of CBCT images acquired with a 0.2-mm voxel size may have facilitated the detection of small accessory canals and subtle anatomical variations that

		Frequency (n)	Percent (%)
Gender	Female	533	67.1
	Male	261	32.9
Age groups	10-20 years	68	8.6
	21-30 years	94	11.8
	31-40 years	157	19.8
	41-50 years	206	25.9
	51-60 years	155	19.5
	61-70 years	96	12.1
	71-80 years	18	2.3
	No variation	108	13.6
Classification of variations	Type 1	23	2.9
	Type 2	228	28.7
	Type 3	367	46.2
	Type 4	68	8.6
	Total edentulism	60	7.6
	No molar missing	261	32.9
Classification of edentulism	One missing molar	158	19.9
	Two missing molars	170	21.4
	Three missing molars	145	18.3
Side of assessment	Right side	409	51.5
	Left side	385	48.5

Table 2. The relationship of classifications of variations with gender, age groups, dentition status, and sides

		Classification of variations					p-value
		No variation	Type 1	Type 2	Type 3	Type 4	
Gender	Female	78	15	140	258	42	0.123
	Male	30	8	88	109	26	
Age groups	10-20 years	10	0	47	10	1	0.001*
	21-30 years	7	2	50	31	4	
	31-40 years	18	5	57	62	15	
	41-50 years	26	8	48	99	25	
	51-60 years	24	3	20	91	17	
	61-70 years	20	5	4	62	5	
	71-80 years	3	0	2	12	1	
Classification of edentulism	Total edentulism	0	3	0	50	7	0.001*
	No molar missing	108	1	108	36	8	
	One missing molar	0	9	80	57	12	
	Two missing molars	0	5	40	100	25	
	Three missing molars	0	5	0	124	16	
Side of assessment	Right side	43	11	124	196	35	0.120
	Left side	65	12	104	171	33	

*: p<0.05, Pearson chi-square

Table 3. Distribution of mandibular canal variation subtypes according to the Naitoh classification

Naitoh subtype	Description	n	%
Type 1A	Intersecting forward canal	5	0.7
Type 1B	Non-intersecting forward canal	18	2.6
Type 2C	Branch reaching the first molar	32	4.7
Type 2D	Branch reaching the second molar	77	11.2
Type 2E	Branch reaching the third molar	119	17.3
Type 3F	Retromolar canal	367	53.5
Type 4G	Branch diverging from the buccal wall of the main canal	60	8.7
Type 4H	Branch diverging from the lingual wall of the main canal	8	1.2
Total		686	100.0

Percentages were calculated based on the total number of detected mandibular canal variations (n=686)

might be overlooked at larger voxel sizes. In addition, the use of a consensus-based final evaluation may have reduced the likelihood of missed findings and contributed to the identification of a greater number of mandibular canal variations. Therefore, differences in prevalence rates among studies should be interpreted not only in terms of population characteristics but also with consideration of imaging parameters and assessment protocols.

In the present study, the most frequently observed mandibular canal variation was Type 3 (retromolar canal), with a prevalence of 46.2%. This finding partially differs from some previously published studies. Naitoh et al. (16) reported that the most common mandibular canal variation was the anterior canal type (55.3%). In contrast, Kang et al. (21) found that the retromolar canal was the most

common variation type (52.5%) in a Korean population. Özarıslantürk and Avsever (19) also reported that the retromolar canal was the most frequently observed variation (63.7%). However, studies reporting lower frequencies are also available. Bilecenoglu and Tuncer (22) identified retromolar canals in 25% of their Turkish sample, whereas the lowest prevalence was reported by Kuribayashi et al. (15), who detected retromolar canals in only 1.7% of 252 patients evaluated with CBCT. In a CBCT-based study conducted in the Çukurova subpopulation, Tumanı et al. (23) reported the anterior canal as the most common variation type (31.4%), followed by the dental canal (29.7%), retromolar canal (26.3%), and buccolingual canal (12.7%). These findings differ from the results of the present study, in which the retromolar canal was the predominant variation type. These differences in

reported prevalence rates suggest that mandibular canal morphology may vary among populations. Among the evaluated subtypes, Type 3F (retromolar canal) was the most prevalent subtype, whereas Type 1A (intersecting forward canal) and Type 4H (lingual buccolingual canal) were the least prevalent subtypes. Within the dental canal group, Type 2E (branch reaching the third molar) was observed more frequently than Type 2D and Type 2C. Similarly, Özarslantürk and Avsever (19) found retromolar canal-related variations to be the most common (63.7%), followed by dental canal variations (30%), whereas forward canal variations accounted for only 5% of cases. In contrast, Orhan et al. (9) reported the forward canal as the most common variation type (20.4%), with non-intersecting forward canals (13.8%) more frequent than intersecting ones. Retromolar canals were observed in 20% of cases, while dental canals were less common overall (4.4%), with the third molar branch (2.7%) representing the most common dental subtype. Buccolingual canals were found in 8.8% of cases, with buccal branches (5.7%) slightly more frequent than lingual branches (3.1%). These differences in subtype distribution may reflect methodological and population-related variations.

In the present study, no statistically significant difference was found between sexes regarding the distribution of mandibular canal variations. This finding is consistent with several previous studies reported in the literature. Orhan et al. (9), in a CBCT-based study conducted in a Turkish population, detected mandibular canal variations in 46.5% of 242 patients and reported that these variations did not differ significantly between males and females. Similarly, Özarslantürk and Avsever (19) also reported that mandibular canal variations were not associated with sex. Similarly, Tumani et al. (23) also found no significant association between bifid mandibular canal distribution and sex. However, in a systematic review and meta-analysis conducted by Asghar et al. (24), mandibular canal variations were reported to be 25% more prevalent in males than in females. Taken together, these findings suggest that the influence of sex on mandibular canal variations remains inconclusive, and further large-scale studies are warranted.

When the relationship between age groups and mandibular canal variations was evaluated in the present study, Type 3 variation was found to be significantly more frequent in the 41-50 years age group ($p=0.001$). Although this

finding indicates a potential association between age and mandibular canal variations, the independent effect of age cannot be determined because potential confounding factors, such as edentulism and dental status, were not controlled for in the analysis. Studies investigating the effect of age on mandibular canal variations are limited. Özarslantürk and Avsever (19) reported no significant differences among age groups. Likewise, in the comprehensive systematic review and meta-analysis by Asghar et al. (24), no statistically significant association was found between age and the prevalence of mandibular canal variations. Further studies incorporating multivariable analyses and larger populations are needed to clarify the relationship between age and mandibular canal variations.

In the present study, when the relationship between edentulism and mandibular canal variations was evaluated, the frequency of Type 3 variation was found to be significantly increased particularly in individuals with the absence of one or two molar teeth. This finding indicates a possible association between posterior tooth loss and mandibular canal variations; however, the underlying mechanism remains unclear. Similarly, Bilecenoglu and Tuncer (22) reported that alveolar resorption resulting from edentulism may alter the course of the canal and render existing variations more radiographically apparent. Therefore, careful evaluation of mandibular canal variations before surgical procedures is particularly important in patients with posterior tooth loss. Von Arx and Bornstein (13) also suggested that the increased radiographic visibility of tubular structures with age or edentulism may not necessarily reflect true anatomical changes.

Regarding side-related (right/left) distribution of variations, no statistically significant difference was observed in the present study. Similar findings were reported by Tumani et al. (23) and Özarslantürk and Avsever (19), both of whom found no significant differences between the right and left sides. In contrast, in the systematic review and meta-analysis conducted by Asghar et al. (24), the likelihood of detecting mandibular canal variations on the right side was reported to be 21% higher than on the left side. In the same study, unilateral bifid mandibular canal variations were reported to occur approximately three times more frequently than bilateral variations. Further studies are needed to clarify the potential influence of side on mandibular canal variations.

Table 4. Observer agreement for mandibular canal variation classification			
Comparison	Weighted kappa	95% CI	Interpretation
Observer 1 vs. observer 2	0.838	(0.796-0.883)	Very good
Observer 1 vs. final consensus	0.919	(0.881-0.952)	Excellent
Observer 2 vs. final consensus	0.894	(0.864-0.932)	Very good
0.81-1.00 almost perfect (25) CI: Confidence interval			

Study Limitations

Several limitations of this study should be considered when interpreting the findings. First, due to the retrospective cross-sectional design, causal relationships cannot be established. The results only demonstrate associations between the evaluated variables and do not permit cause-and-effect inferences.

Second, although a relatively large sample was evaluated, the study was conducted at a single center and within a Turkish population. Therefore, the generalizability of the findings to other ethnic or geographic populations may be limited.

Third, unilateral and bilateral hemi-mandibles were analyzed as separate observational units. Although this approach increased the sample size, potential correlations between the right and left sides of the same individual may have affected the statistical independence of the observations.

Fourth, mandibular canal variations were classified according to the Naitoh classification system. Although this system is widely used, certain anatomical variations that do not fully conform to these predefined categories may have been overlooked or simplified.

Finally, potential confounding factors such as dental history, degree of alveolar bone resorption, systemic diseases, and craniofacial morphological characteristics were not assessed. Future multicenter prospective studies with standardized protocols and more comprehensive clinical variables would be valuable.

Additionally, one limitation is that the data were acquired from a single CBCT device. The use of one imaging system may introduce potential biases in detection or measurement of variations. Future studies involving multiple imaging devices may provide broader applicability.

Conclusion

This study demonstrated a high prevalence of mandibular canal variations, with Type 3 being the most frequently observed subtype. These findings indicate that mandibular canal anatomy exhibits considerable variability and that such variations should be carefully assessed during clinical evaluation and treatment planning.

Although no significant association was found with sex or side distribution, significant relationships were identified with age and edentulism status. Type 3 variation was more frequently observed in individuals aged 41-50 years and in patients with posterior molar tooth loss.

Therefore, detailed anatomical assessment using CBCT before implant placement, third molar surgery, and other invasive mandibular procedures may contribute to minimizing potential neurovascular complications and improving treatment planning.

Ethics

Ethics Committee Approval: This retrospective cross-sectional investigation was approved by the Non-Interventional Ethics Committee of Bezmiâlem Vakıf University (approval no: E-54022451-050.05.04-108546, decision no: 2023/177, date: 26.05.2023).

Informed Consent: Retrospective study.

Declaration on the Use of Artificial Intelligence (AI)

For transparency, the authors note that an artificial intelligence-assisted language model was utilized solely for 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: E.A., N.K., N.G., Design: E.A., İ.S., Data Collection or Processing: E.A., N.K., N.G., Analysis or Interpretation: E.A., N.K., N.G., İ.S., Literature Search: E.A., N.K., N.G., İ.S., Writing: E.A., N.K., N.G., İ.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. Varvara G, Feragalli B, Turkyilmaz I, D'Alonzo A, Rinaldi F, Bianchi S, et al. Prevalence and characteristics of accessory mandibular canals: a cone-beam computed tomography study in a European adult population. *Diagnostics (Basel)*. 2022;12:1885.
2. Carter RB, Keen EN. The intramandibular course of the inferior alveolar nerve. *J Anat*. 1971;108:433-40.
3. Claes V, Wackens G. Bifid mandibular canal: literature review and case report. *Dentomaxillofac Radiol*. 2005;34:55-8.
4. Rashsuren O, Choi JW, Han WJ, Kim EK. Assessment of bifid and trifid mandibular canals using cone-beam computed tomography. *Imaging Sci Dent*. 2014;44:229-36.
5. Patel S, Dawood A, Ford TP, Whaites E. The potential applications of cone beam computed tomography in the management of endodontic problems. *Int Endod J*. 2007;40:818-30.
6. Misch CE. Density of bone: effect on treatment plans, surgical approach, healing, and progressive bone loading. *Int J Oral Implantol*. 1990;6:23-31.
7. Naitoh M, Hiraiwa Y, Aimiya H, Gotoh M, Aiji Y, Izumi M, et al. Bifid mandibular canal in Japanese. *Implant Dent*. 2007;16:24-32.
8. Scarfe WC, Farman AG, Sukovic P. Clinical applications of cone-beam computed tomography in dental practice. *J Can Dent Assoc*. 2006;72:75-80.
9. Orhan K, Aksoy S, Bilecenoglu B, Sakul BU, Paksoy CS. Evaluation of bifid mandibular canals with cone-beam computed tomography in a Turkish adult population: a retrospective study. *Surg Radiol Anat*. 2011;33:501-7.

10. Chen Y, Liao Q, Wu Y, He R, Zhang N, Liao Y, et al. Study of the mandibular canal and its surrounding canals by multi-view cone-beam computed tomography. *Insights Imaging*. 2024;15:103.
11. Alali Y, Mohammed WA, Alabulkarim M, Alshahrani A, Almawh A. Assessment of bifid mandibular canals using cone beam computed tomography in general population: a retrospective evaluation. *Eur Rev Med Pharmacol Sci*. 2024;28:1741-50.
12. Khojastepour L, Nasiri MM, Hashemi K, Ahrari F. Mandibular canal branching assessed with cone-beam computed tomography. *J Craniofac Surg*. 2023;34:e767-71.
13. Von Arx T, Bornstein MM. The bifid mandibular canal in three-dimensional radiography: morphologic and quantitative characteristics. *SWISS DENTAL JOURNAL SSO-Science and Clinical Topics*. 2021;131:10-28.
14. Mallya S, Lam E. *White and Pharoah's oral radiology: principles and interpretation*: Elsevier Health Sciences; 2018.
15. Kuribayashi A, Watanabe H, Imaizumi A, Tantanapornkul W, Katakami K, Kurabayashi T. Bifid mandibular canals: cone beam computed tomography evaluation. *Dentomaxillofac Radiol*. 2010;39:235-9.
16. Naitoh M, Hiraiwa Y, Aimiya H, Arijji E. Observation of bifid mandibular canal using cone-beam computerized tomography. *Int J Oral Maxillofac Implants*. 2009;24:155-9.
17. Doğan ME, Kotanlı S. Assessment of anatomic variations in the mandible by CBCT. *NEU Dent J*. 2024;6:270-9.
18. Orhan K, Özemre MÖ, Köseoğlu Seçgin C, Gülşahi A. Alt anterior dişlerin kök kanal morfolojisinin konik ışınli bilgisayarlı tomografi kullanılarak değerlendirilmesi [Evaluation of root canal morphology of mandibular anterior teeth using cone beam computed tomography]. *Türkiye Klinikleri J Dental Sci*. 2018;24:190-6. Turkish.
19. Özarslantürk S, Avsever İH. Türk toplumunda mandibular kanal varyasyonlarının konik ışınli bilgisayarlı tomografi ve ortopantomografi yöntemleri ile karşılaştırmalı olarak incelenmesi [The investigation of mandibular canal variations in the turkish society comparative with cone beam computerized tomography and orthopantomography methods]. *Life Med Sci*. 2023;2:70-85. Turkish.
20. Samieirad S, Aryana M, Mazandarani A, Misagh Toupkanloo I, Eidi M, Moqarabzadeh V, et al. Prevalence of bifid mandibular canal: a systematic review and meta-analysis. *World J Plast Surg*. 2023;12:11-9.
21. Kang JH, Lee KS, Oh MG, Choi HY, Lee SR, Oh SH, et al. The incidence and configuration of the bifid mandibular canal in Koreans by using cone-beam computed tomography. *Imaging Sci Dent*. 2014;44:53-60.
22. Bilecenoglu B, Tuncer N. Clinical and anatomical study of retromolar foramen and canal. *J Oral Maxillofac Surg*. 2006;64:1493-7.
23. Tumanı B, Duyan H, Evlice B. Çukurova popülasyonunda BIFID mandibular kanal prevalansı: BIR KİBT çalışması [Prevalance of BIFID mandibular canal in Cukurova subpopulation: a CBCT study. *Sağlık Bilimleri Dergisi*. 2022;31:287-93.
24. Asghar A, Priya A, Ravi KS, Iwanaga J, Tubbs RS, Naaz S, et al. An evaluation of mandibular canal variations: a systematic review and meta-analysis. *Anat Sci Int*. 2023;98:176-84.
25. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-74.



Establishment of Age-specific Reference Intervals for Serum Magnesium Concentrations in Healthy Children Aged 0-18 Years: A Comparative Analysis of Non-parametric, Bootstrap, and Gaussian Mixture Model Approaches

0-18 Yaş Arası Sağlıklı Çocuklarda Serum Magnezyum Düzeylerinin Yaşa Özgü Referans Aralıklarının Belirlenmesi: Non-parametrik, Bootstrap ve Gaussian Mixture Model Yaklaşımlarının Karşılaştırmalı Analizi

¹ Adile ÖZERTÜRK¹, ² Fatma KUMBASAR¹, ³ Mehmet Murat YEKREK¹, ² Özge PASİN², ³ Eray Metin GÜLER^{1,3}

¹University of Health Sciences Türkiye, Hamidiye Faculty of Medicine, Haydarpaşa Numune Health Application and Research Center, Department of Medical Biochemistry, İstanbul, Türkiye

²University of Health Sciences Türkiye, Hamidiye Faculty of Medicine, Department of Biostatistics and Medical Informatics, İstanbul, Türkiye

³University of Health Sciences Türkiye, Hamidiye Faculty of Medicine, Department of Medical Biochemistry, İstanbul, Türkiye

ABSTRACT

Objective: This study aimed to establish age-specific reference intervals (RIs) for serum magnesium (Mg) concentrations in healthy children aged 0-18 years and to compare the derived intervals with manufacturer-provided reference ranges.

Methods: This retrospective, cross-sectional study included 1,725 healthy children who presented to the Medical Biochemistry Laboratory of University of Health Sciences Türkiye, Haydarpaşa Numune SUAM between November 1, 2022, and February 18, 2023. Serum Mg was measured using the colorimetric end-point method on the Roche Diagnostics COBAS 8000 (c702 module). RIs were calculated using the non-parametric 2.5th-97.5th percentile method according to CLSI EP28-A3c and IFCC guidelines, and validated with bootstrap resampling (95% confidence interval) and Gaussian mixture model (GMM) analysis.

ÖZ

Amaç: Bu çalışma, 0-18 yaş arası sağlıklı pediatrik bireylerde serum magnezyum (Mg) düzeyleri için yaşa özgü referans aralıklarını belirlemeyi ve elde edilen sonuçları üretici firma referanslarıyla karşılaştırmayı amaçlamıştır.

Yöntemler: Retrospektif ve kesitsel olarak tasarlanan bu çalışmada, 1 Kasım 2022 ve 18 Şubat 2023 tarihleri arasında Sağlık Bilimleri Üniversitesi, Haydarpaşa Numune SUAM Tıbbi Biyokimya Laboratuvarı'na başvuran 1,725 sağlıklı çocuk değerlendirilmiştir. Serum Mg düzeyleri Roche Diagnostics COBAS 8000 (c702) analizöründe kolorimetrik end-point yöntemle ölçülmüştür. Referans aralıkları, CLSI EP28-A3c ve IFCC önerileri doğrultusunda non-parametrik yüzde 2,5.-97,5. yöntemiyle hesaplanmış; bootstrap (%95 güven aralığı) ve Gaussian mixture model (GMM) analizleriyle doğrulanmıştır.

Address for Correspondence: Assoc. Prof., Eray Metin Güler, University of Health Sciences Türkiye, Hamidiye Faculty of Medicine Department of Medical Biochemistry, İstanbul, Türkiye

E-mail: eraymetinguler@gmail.com

ORCID IDs of the authors: A.Ö.: 0009-0005-9126-8540, F.K.: 0009-0005-0756-2692, M.M.Y.: 0000-0002-0880-8421, Ö.P.: 0000-0001-6530-0942, E.M.G.: 0000-0003-4351-1719

Cite this article as: Özertürk A, Kumbasar F, Yekrek MM, Pasin Ö, Güler EM. Establishment of age-specific reference intervals for serum magnesium concentrations in healthy children aged 0-18 years: a comparative analysis of non-parametric, bootstrap, and Gaussian mixture model approaches. Bezmialem Science. 2026;14(4):357-364

Received: 14.11.2025

Accepted: 29.06.2026

Published date: 26.10.2026



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

ABSTRACT

Results: Serum Mg concentrations differed significantly among age groups (Kruskal-Wallis, $p<0.05$). The highest median level was found in the 7-12-year group (2.21 mg/dL), and the lowest in the 13-18-year group (2.06 mg/dL). Serum magnesium concentrations increased from early to middle childhood, reaching their highest median levels at 7-12 years, followed by a decline during adolescence (13-18 years). The GMM-derived reference limits were generally comparable to those obtained using the non-parametric method. The derived age-specific reference intervals generally exhibited higher upper reference limits than the age-dependent expected values reported in the manufacturer's documentation.

Conclusion: This study provides locally derived, age-specific serum magnesium reference intervals for a Turkish pediatric population, contributing to the available evidence on population-specific reference intervals in pediatric laboratory medicine. The combined application of non-parametric, bootstrap, and GMM methods enhances methodological robustness in RI establishment. These findings emphasize the necessity of population-specific RIs and offer a foundation for developing a national pediatric RI database in Türkiye.

Keywords: Magnesium, pediatric reference intervals, non-parametric method, Gaussian mixture model, bootstrap analysis, clinical biochemistry

Öz

Bulgular: Serum Mg düzeyleri yaş grupları arasında anlamlı farklılık göstermiştir (Kruskal-Wallis, $p<0,05$). En yüksek ortalama düzey 7-12 yaş grubunda (2,21 mg/dL), en düşük ortalama düzey ise 13-18 yaş grubunda (2,06 mg/dL) saptanmıştır. Serum magnezyum düzeyleri erken çocukluktan orta çocukluk dönemine kadar artış göstererek 7-12 yaş grubunda en yüksek ortalama düzeye ulaşmış, ergenlik döneminde (13-18 yaş) ise azalmıştır. GMM ile elde edilen referans sınırları, non-parametrik yöntemle hesaplanan sınırlarla genel olarak benzerlik göstermiştir. Elde edilen yaşa özgü referans aralıklarının üst sınırları, üretici dokümanında bildirilen yaşa bağlı beklenen değerlerin üst sınırlarından genel olarak daha yüksek bulunmuştur.

Sonuç: Bu çalışma, Türkiye'deki tek merkezli bir pediatrik popülasyonda 0-18 yaş grubuna özgü serum magnezyum referans aralıklarını belirleyerek pediatrik laboratuvar tıbbında popülasyona özgü referans aralıklarına ilişkin mevcut bilimsel verilere katkı sağlamıştır. Non-parametrik, bootstrap ve GMM yaklaşımlarının birlikte kullanımı, referans aralığı belirlemede metodolojik güvenilirliği artırmaktadır. Bulgular, ulusal düzeyde pediatrik referans aralık veri tabanı oluşturulması için önemli bir temel sunmaktadır.

Anahtar Kelimeler: Magnezyum, pediatrik referans aralıkları, non-parametrik yöntem, Gaussian mixture model, bootstrap analizi, klinik biyokimya

Introduction

Magnesium (Mg) is an essential macromineral required for the maintenance of vital physiological functions and represents the second most abundant intracellular cation after potassium. Approximately 60% of the body's total Mg content is stored within bone tissue, 20% within skeletal muscle, and the remaining fraction in soft tissues such as the liver, heart, and kidneys (1). Only a small fraction of total body magnesium is present in the circulation, with approximately 55-70% of serum magnesium existing in the free ionized form, 20-30% bound to proteins (predominantly albumin), and the remainder complexed with various anions (2).

Mg acts as a mandatory cofactor for more than 300 enzymatic reactions and is intimately involved in virtually all major metabolic pathways. It plays a central role in adenosine triphosphate (ATP) synthesis and energy metabolism, nucleic acid synthesis, protein translation, glucose homeostasis, muscle contraction, cardiac excitability, neuromuscular transmission, and regulation of vascular tone (3,4). Consequently, disturbances in Mg homeostasis have been associated with a wide range of clinical conditions, including cardiovascular diseases, metabolic syndrome, type 2 diabetes, neuromuscular disorders, arrhythmias, and hypertension (1,5).

In routine laboratory practice, serum total Mg concentration remains the most commonly used biomarker. Although

serum Mg represents only a small fraction of total body Mg stores, it is widely utilized because of its practicality and applicability in clinical decision-making (6). However, serum Mg levels are influenced by numerous biological and analytical variables—including age, sex, dietary intake, hormonal status, pregnancy, comorbid conditions, medication use, and methodological differences—resulting in significant inter-laboratory variability. This heterogeneity underscores the necessity of establishing locally validated reference intervals (RIs) (7,8).

Historically, adult RIs for serum Mg have been reported as 0.75-0.95 mmol/L (1.82-2.31 mg/dL), largely based on NHANES-I data obtained in the 1970s (6). Emerging evidence has suggested that serum Mg concentrations below approximately 0.85 mmol/L may be associated with increased clinical risk or suboptimal Mg status (2,9). However, this proposed value should be interpreted as a clinical risk or adequacy threshold rather than as a statistically derived RI lower limit. In contrast, RI boundaries are established from the distribution of measurements in a defined healthy population, typically using the 2.5th and 97.5th percentiles according to Clinical and Laboratory Standards Institute (CLSI)/International Federation of Clinical Chemistry (IFCC) recommendations. In pediatric populations, Mg levels differ substantially from those in adults due to factors such as rapid somatic growth, endocrinological variation, and ongoing organ maturation. As a result, age-appropriate and population-specific RIs are essential for accurate clinical interpretation (10,11).

To address this need, the IFCC and the CLSI recommend that each laboratory establish its own RIs specific to its population, analytical platforms, and measurement procedures, as outlined in the EP28-A3c guideline (12). This recommendation is particularly critical for pediatric RI studies, where age, sex, biological variation, and analytical uncertainty must be carefully considered. Supporting this approach, the Canadian Laboratory Initiative on Pediatric Reference Intervals (CALIPER) project has demonstrated significant age-dependent variability in numerous biochemical parameters, including Mg (10). However, despite the growing recognition of population-specific pediatric RIs, nationally validated serum Mg RI studies in the Turkish pediatric population remain limited. Most laboratories in Türkiye continue to rely on manufacturer-provided intervals or internationally derived datasets, which may not adequately reflect local demographic, ethnic, nutritional, and analytical characteristics. Therefore, establishing locally derived pediatric serum Mg RIs may contribute to improving laboratory interpretation and clinical decision-making in Turkish pediatric populations.

Given the methodological heterogeneity across laboratories and the physiological variability in pediatric age groups, the establishment of locally derived, method-specific RIs is of paramount importance. In this context, the present study aims to define age-specific RIs for serum Mg concentrations in healthy children aged 0-18 years who presented to the Medical Biochemistry Clinic of University of Health Sciences Türkiye, Haydarpaşa Numune SUAM, thereby contributing novel and clinically meaningful data to pediatric laboratory medicine.

Methods

Study Design and Ethical Approval

This study was designed as a retrospective and cross-sectional descriptive analysis. The objective was to establish age-specific RIs for serum Mg concentrations in pediatric individuals admitted to the Medical Biochemistry Clinic of University of Health Sciences Türkiye, Haydarpaşa Numune SUAM. All procedures were performed in accordance with the CLSI EP28-A3c guideline and the recommendations of the IFCC.

Ethical approval was obtained from the Hamidiye Scientific Research Ethics Committee of the University of Health Sciences Türkiye (decision no: 2025-4/12, date: 27.02.2025). Because the study was retrospective and based exclusively on anonymized laboratory and hospital information system data, the requirement for individual informed consent was waived by the relevant ethics committee.

The study adhered to the ethical principles of the Declaration of Helsinki (2013 revision), and confidentiality of personal data was strictly maintained.

Study Population

The dataset consisted of routine biochemical test results from pediatric patients aged 0-18 years (inclusive) who

presented to the laboratory between November 1, 2022, and February 18, 2023. Age was calculated in completed years, and all participants aged 0 through 18 years were eligible for inclusion. Data were retrieved from the Laboratory Information System (LIS) using Structured Query Language queries, including age, sex, admission date, and serum Mg results. Only the first measurement per individual was included.

Clinically healthy individuals were included. Subjects with acute infection, chronic disease, metabolic or hormonal disorders, systemic illness, intensive care follow-up, or chronic medication use were excluded. Samples showing hemolysis, lipemia, or icterus interference were rejected. Individuals with pathological albumin, calcium, or creatinine levels were also excluded. Records that were incomplete, erroneous, or duplicated were removed.

Clinical health status was assessed using available hospital information system records, International Classification of Diseases (ICD)-coded diagnoses, clinical admission data, and laboratory-based exclusion criteria. Although the retrospective design precluded direct clinical examination, individuals with documented acute or chronic systemic diseases, renal, hepatic, endocrine, metabolic, or inflammatory disorders were excluded from the reference population.

A total of 1,725 pediatric samples met the inclusion criteria. Each age-specific subgroup included more than 120 individuals (0-6 years: n=512; 7-12 years: n=641; 13-18 years: n=572), satisfying the minimum sample size recommendation for the non-parametric estimation of reference intervals according to CLSI EP28-A3c guidelines.

Sample Collection and Biochemical Measurements

Blood samples were collected from the antecubital vein between 08:00 and 10:00 into serum separator tubes containing clot activator. After clot formation, samples were centrifuged at 3000×g for 10 minutes. Serum magnesium concentrations were measured using a colorimetric end-point assay based on the formation of a magnesium-xylidyl blue complex on the cobas 8000 c702 analyzer (Roche Diagnostics, Mannheim, Germany). Measurements were performed according to the manufacturer's reagent instructions, and results were expressed in mg/dL (13). Internal quality control was performed using two levels of PreciControl ClinChem Multi control materials (Roche Diagnostics GmbH, Mannheim, Germany). External quality assessment was conducted through the Randox International Quality Assessment Scheme (RIQAS; Randox Laboratories Ltd., Crumlin, United Kingdom) (14,15). The analytical coefficient of variation (CVa) for serum magnesium measurements was 1.38%, indicating low analytical imprecision under the laboratory's routine quality control conditions.

Data Processing and Quality Control

During preprocessing, incomplete or inconsistent records were excluded. Potential outliers were identified using the interquartile range (IQR) method, with lower and upper fences defined as $Q1 - 1.5 \times IQR$ and $Q3 + 1.5 \times IQR$, respectively. The application of these criteria and the number of observations excluded from each age group should be reported to clarify their potential impact on the estimated reference limits. Distribution characteristics were evaluated using histograms and boxplots, and normality was assessed via the Kolmogorov-Smirnov test.

Age Group Stratification

Because pediatric Mg levels vary with growth, hormonal maturation, and developmental physiology, the population was divided into three age groups:

- 0-6 years: early childhood,
- 7-12 years: middle childhood,
- 13-18 years: adolescence.

Age-specific partitioning was evaluated using the Harris-Boyd and Lahti approaches. Sex-specific partitioning within the predefined age groups was assessed according to CLSI/IFCC recommendations using the Harris-Boyd criteria. Although minor differences in serum magnesium concentrations were observed between males and females during adolescence, separate sex-specific reference intervals were not established because the available findings did not provide sufficient evidence to support further partitioning.

Determination of RIs

RIs were calculated using non-parametric methods according to CLSI EP28-A3c and IFCC guidelines. The 2.5th and 97.5th percentiles were estimated for each age group. Bootstrap resampling was used to estimate 95% confidence intervals (CIs) for the lower and upper non-parametric reference limits within each age group. Gaussian mixture model (GMM) analysis was performed to evaluate potential multimodality in serum Mg distributions. The optimal number of mixture components for each age group was determined using the Bayesian information criterion (BIC), and the model with the lowest BIC value was selected as the best-fitting model. GMM analysis was performed to explore the distributional characteristics of serum magnesium concentrations within each age group. The number of mixture components was selected using the BIC. The resulting model-based reference limits were descriptively compared with the corresponding non-parametric reference limits. When required, Box-Cox transformation improved distribution symmetry.

Quantile regression analysis was performed to evaluate continuous age-related changes in serum Mg concentrations across the 2.5th, 50th, and 97.5th percentiles without assuming Gaussian distribution. Quantile regression coefficients and

pseudo-R² values were estimated to characterize age-related changes in serum magnesium concentrations at the 2.5th, 50th, and 97.5th percentiles. The estimated coefficients and model fit statistics are presented in Supplementary Table 1 (16,17).

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 22.0 (IBM Corp., Armonk, NY, USA), Python version 3.11 (Python Software Foundation, Wilmington, DE, USA), and R version 4.4.2 (R Foundation for Statistical Computing, Vienna, Austria). One-way analysis of variance (ANOVA) with Tukey post-hoc testing was used for normally distributed variables; Kruskal-Wallis and Dunn tests were applied otherwise. Bonferroni correction was applied to account for multiple pairwise comparisons. Effect sizes were assessed using eta-squared (η^2) for analysis of variance and the rank-based effect size (r) for non-parametric pairwise comparisons. A p-value <0.05 indicated statistical significance. RIs obtained in this study were compared with manufacturer-provided pediatric values (Roche Diagnostics).

Ethical Considerations

All analyses were performed retrospectively using anonymized data. The study fully complied with the ethical standards of the Declaration of Helsinki (2013 revision).

Results

A total of 1,725 healthy pediatric individuals who met the inclusion criteria and presented to the Medical Biochemistry Clinic of University of Health Sciences Türkiye, Haydarpaşa Numune SUAM between November 1, 2022, and February 18, 2023, were included in the analysis. The mean age of participants was 9.8 ± 4.6 years, with a sex distribution of 52.4% male and 47.6% female.

Serum Mg concentrations exhibited a right-skewed distribution, and the Kolmogorov-Smirnov test indicated a statistically significant departure from normality ($p < 0.05$). Given the distributional characteristics of the data, age-specific reference intervals were estimated using the non-parametric 2.5th and 97.5th percentile method in accordance with CLSI EP28-A3c recommendations.

Serum Mg Levels by Age Group

Participants were stratified into three age groups based on developmental physiology:

- Group 1: 0-6 years ($n=512$),
- Group 2: 7-12 years ($n=641$),
- Group 3: 13-18 years ($n=572$).

The Kruskal-Wallis test indicated an overall statistically significant difference in serum magnesium concentrations among the three age groups ($p < 0.05$). Descriptively, the highest median serum magnesium concentration was

Table 1. Age-specific pediatric serum magnesium reference intervals determined using non-parametric percentiles with 95% bootstrap confidence intervals (CIs) for the lower and upper reference limits

Age group	n	Lower limit (2.5 th) (CI)	Upper limit (97.5 th) (CI)	Median
0-6	512	1.79 (1.76-1.82)	2.42 (2.40-2.47)	2.15
7-12	641	1.82 (1.81-1.84)	2.48 (2.45-2.51)	2.21
13-18	572	1.74 (1.71-1.76)	2.37 (2.35-2.39)	2.06

Reference intervals were calculated using the non-parametric 2.5th-97.5th percentile method. Values in brackets indicate 95% bootstrap CIs for the lower and upper reference limits

observed in the 7-12-year age group (2.21 mg/dL), followed by the 0-6-year group (2.15 mg/dL), whereas the lowest median was observed in the 13-18-year group (2.06 mg/dL). These findings indicate an age-related variation in serum magnesium concentrations, although the overall Kruskal-Wallis result does not establish which individual age groups differ significantly.

RI Determination

Following CLSI EP28-A3c and IFCC recommendations, RIs were established using the non-parametric 2.5th-97.5th percentile method. The lower and upper reference limits, their corresponding 95% bootstrap CIs, and median values are presented in Table 1.

GMM and Bootstrap Analysis

GMM analysis was performed to examine the distributional characteristics of serum magnesium concentrations within each age group. The resulting GMM-derived reference limits are presented alongside the non-parametric reference limits in Table 2. GMM-derived reference limits were generally comparable to those obtained using the non-parametric method, with relatively small differences in the estimated lower and upper reference limits across age groups.

Quantile regression analysis was used to evaluate age-related changes in serum magnesium concentrations across the 2.5th, 50th, and 97.5th percentiles. The estimated trajectories suggested relatively stable magnesium concentrations during early childhood, followed by lower concentrations during adolescence. These patterns should be interpreted in conjunction with the estimated regression coefficients and their associated statistical uncertainty.

Figure 1 presents the quantile regression curves for serum Mg concentrations according to age. The 2.5th, 50th, and 97.5th percentile trajectories demonstrate relatively stable Mg levels during early childhood, followed by lower serum Mg concentrations during adolescence.

Comparison with Manufacturer RIs

The Roche Diagnostics Magnesium Gen.2 reagent documentation provides age-dependent expected values for serum magnesium, including 1.7-2.3 mg/dL for children aged 5 months to 6 years, 1.7-2.1 mg/dL for those aged 6-12 years, and 1.7-2.2 mg/dL for those aged 12-20 years

(13). In the present study, the upper reference limits were 2.42 mg/dL, 2.48 mg/dL, and 2.37 mg/dL for the 0-6, 7-12, and 13-18-year age groups, respectively. These limits were higher than the upper expected values reported in the manufacturer's documentation for the corresponding pediatric age ranges. However, because the manufacturer's age categories do not fully align with those used in the present study, direct age-matched comparisons were not possible. The observed differences may reflect population-specific biological variation, differences in reference population selection, and preanalytical factors. Therefore, these findings should be interpreted descriptively, and the relative contributions of these factors cannot be determined from the present data.

Overall Assessment

Serum Mg levels showed marked variability across pediatric developmental stages. The decline during adolescence may reflect increased growth velocity, hormonal changes, and dietary variations. The observed differences between manufacturer-provided and locally derived reference intervals highlight the importance of evaluating population-specific reference intervals in clinical laboratories. This study provides age-specific local RIs for serum Mg in healthy children aged 0-18 years and demonstrates the complementary value of non-parametric, bootstrap, and GMM methods in RI studies.

Discussion

In this study, age-specific RIs for serum Mg concentrations in healthy children aged 0-18 years were established using non-parametric, bootstrap, and GMM approaches in accordance with CLSI EP28-A3c and IFCC recommendations and were subsequently compared with manufacturer-provided intervals (8). Our findings demonstrate significant age-related differences in serum Mg concentrations, with the highest median concentrations observed during middle childhood and relatively lower values during adolescence. Given the central role of Mg in ATP-dependent reactions, nucleic acid and protein synthesis, muscle contraction, neuromuscular transmission, and intracellular signaling, reliable pediatric RIs are essential for the accurate identification and clinical interpretation of hypo- and hypermagnesemia (1,3,4).

The observed age-dependent differences are biologically plausible. Relatively higher serum Mg concentrations

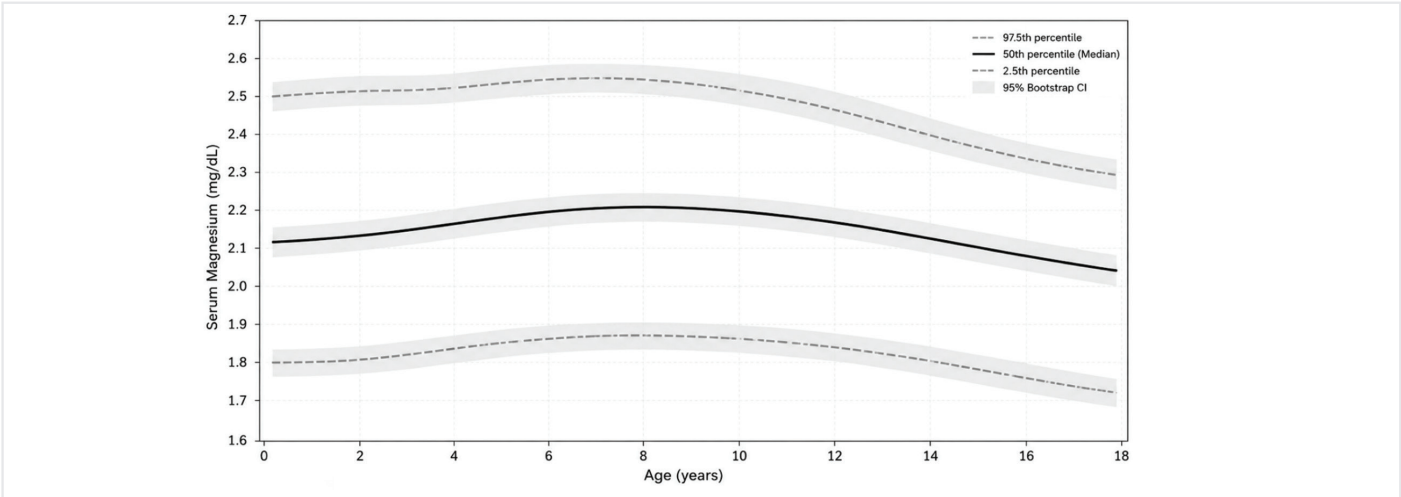


Figure 1. Estimated age-related trajectories of serum magnesium concentrations at the 2.5th, 50th (median), and 97.5th percentiles. The curves suggest an increase in median magnesium concentrations during childhood, followed by a decline during adolescence. Shaded regions represent the reported 95% bootstrap confidence bands around the estimated quantile trajectories
CI: Confidence interval

Table 2. Descriptive comparison of non-parametric and Gaussian mixture model-derived serum magnesium reference limits across pediatric age groups

Age group	Method	Lower limit	Upper limit	Mean ± SD	Correlation (r)
0-6	Non-parametric	1.79	2.42	2.15±0.18	-
	GMM	1.78	2.44	2.16±0.17	0.95
7-12	Non-parametric	1.82	2.48	2.21±0.19	-
	GMM	1.80	2.47	2.22±0.18	0.94
13-18	Non-parametric	1.74	2.37	2.06±0.20	-
	GMM	1.72	2.36	2.05±0.19	0.93

Differences between non-parametric and GMM-derived reference limits are presented descriptively. Formal equivalence or agreement between the two estimation approaches was not established
SD: Standard deviation.

during early and middle childhood may be associated with active bone mineralization, protein synthesis, and increased energy requirements. In contrast, the relatively lower concentrations observed during adolescence may reflect pubertal hormonal changes, rapid somatic growth, sex-related physiological differences, changing dietary patterns, and increased metabolic utilization. Consistent with our findings, data from the CALIPER initiative have demonstrated substantial age- and sex-dependent variability across numerous pediatric biochemical analytes, including electrolytes and minerals (10,11).

From a methodological perspective, the non-parametric 2.5th-97.5th percentile approach was appropriate because serum Mg concentrations deviated significantly from a Gaussian distribution. Bootstrap resampling provided 95% confidence intervals for the estimated lower and upper reference limits, allowing an assessment of their sampling uncertainty. However, these confidence intervals should be interpreted as measures of statistical precision rather than as evidence of the clinical validity or external generalizability of the derived reference intervals (18). Furthermore, the

GMM-derived reference limits were generally comparable to those obtained using the non-parametric approach. However, agreement between reference limits does not independently establish unimodality, which should be evaluated based on the optimal number of mixture components selected by the BIC. The parallel application of non-parametric, bootstrap, and GMM approaches therefore strengthened the methodological reliability of the derived RIs.

Compared with the age-dependent expected values reported in the Roche Diagnostics Magnesium Gen.2 documentation, the reference intervals derived in the present study generally exhibited higher upper reference limits. Nevertheless, differences in age-group definitions and reference population characteristics should be considered when interpreting these findings. These differences may be attributable to population-specific biological characteristics, nutritional and demographic factors, preanalytical conditions, and analytical differences related to the measurement system, reagents, and calibration procedures (7). The need for age-based

partitioning, as recommended by the Harris-Boyd and Lahti approaches, was further supported by both the visual and statistical findings of the quantile regression analysis (19,20).

The principal strengths of the present study include its large sample size ($n=1,725$), low analytical imprecision ($CVa=1.38\%$), verification of analytical performance through two-level internal quality control and external quality assessment, and the simultaneous use of complementary statistical methods. Direct comparison with manufacturer-provided intervals also enhances the clinical relevance of the findings. In addition, the use of LIS-based real-world data reflects routine clinical laboratory practice and increases the practical applicability of the proposed RIs.

Collectively, these findings support the use of locally derived, age-specific, population-specific, and method-specific RIs to improve the interpretation of serum Mg measurements and clinical decision-making in pediatric laboratory medicine.

Study Limitations

The present study has several limitations. First, its retrospective design precluded the direct clinical examination of participants and limited the availability of detailed information regarding several preanalytical and biological variables that may influence serum Mg concentrations, including fasting status, dietary Mg intake, hydration status, and physical activity. Although individuals with abnormal serum albumin concentrations were excluded, variations in albumin levels within the reference range may still have influenced total serum magnesium concentrations, as approximately 20-30% of circulating magnesium is protein-bound, predominantly to albumin.

Second, the study was conducted at a single tertiary-care center, which may limit the generalizability of the findings to the broader pediatric population. Referral-based hospital populations may include a greater proportion of individuals with subclinical or early-stage conditions than community-based cohorts, potentially introducing selection bias into the derived RIs. Nevertheless, extensive laboratory-based exclusion criteria, ICD-coded diagnoses, clinical admission records, and available laboratory findings were reviewed to minimize this potential effect.

Third, dietary, socioeconomic, regional, and ethnic factors could not be comprehensively evaluated. Moreover, although sex-specific partitioning was assessed, the number of participants within narrower age- and sex-specific subgroups may not have been sufficient to identify subtle sex-related differences, particularly during puberty. Therefore, prospective multicenter studies involving geographically, socioeconomically, and demographically diverse pediatric populations are required to externally validate the proposed intervals and support the

development of nationally representative pediatric serum Mg RIs for Türkiye.

Conclusion

This study establishes age-specific serum Mg RIs for healthy children aged 0-18 years using complementary non-parametric, bootstrap, and GMM approaches. The observed age-dependent pattern, characterized by the highest median concentrations during middle childhood and relatively lower concentrations during adolescence, highlights the importance of developmental stage-specific interpretation. The differences between the locally derived RIs and the manufacturer-provided values further demonstrate the clinical importance of population- and method-specific RIs. These findings provide locally derived RI data for a Turkish pediatric population and may support future multicenter studies and the development of broader pediatric RI datasets in Türkiye.

Ethics

Ethics Committee Approval: Ethical approval was obtained from the Hamidiye Scientific Research Ethics Committee of the University of Health Sciences Türkiye (decision 2025-4/12, date: 27.02.2025).

Informed Consent: Given the retrospective design of the study and the use of anonymized data obtained from laboratory and hospital information systems, the requirement for individual informed consent was waived by the Hamidiye Scientific Research Ethics Committee of the University of Health Sciences Türkiye.

Acknowledgements

Confidentiality of Data

All patient-related data were anonymized before analysis, and the confidentiality of personal information was maintained throughout the study.

Data Availability

The datasets generated and analyzed during the present study are available from the corresponding author upon reasonable request and subject to institutional and ethical restrictions concerning patient confidentiality.

Footnotes

Authorship Contributions

Concept: A.Ö., F.K., M.M.Y., E.M.G., Design: A.Ö., F.K., M.M.Y., E.M.G., Data Collection or Processing: Ö.P., Analysis or Interpretation: Ö.P., Literature Search: A.Ö., F.K., M.M.Y., E.M.G., Writing: Ö.P., E.M.G.

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 Link: <https://d2v96fxpocvxx.cloudfront.net/e1c28131-4444-460a-9703-9fa751c405df/content-images/e8341c7f-3a43-47df-8a03-169f59e26854.pdf>

References

1. de Baaij JH, Hoenderop JG, Bindels RJ. Magnesium in man: implications for health and disease. *Physiol Rev*. 2015;95:1-46.
2. Costello RB, Elin RJ, Rosanoff A, Wallace TC, Guerrero-Romero F, Hruby A, et al. Perspective: the case for an evidence-based reference interval for serum magnesium: the time has come. *Adv Nutr*. 2016;7:977-93.
3. Gröber U, Schmidt J, Kisters K. Magnesium in prevention and therapy. *Nutrients*. 2015;7:8199-226.
4. Quamme GA. Molecular identification of ancient and modern mammalian magnesium transporters. *Am J Physiol Cell Physiol*. 2010;298:C407-29.
5. DiNicolantonio JJ, O'Keefe JH, Wilson W. Subclinical magnesium deficiency: a principal driver of cardiovascular disease and a public health crisis. *Open Heart*. 2018;5:e000668. Erratum in: *Open Heart*. 2018;5:e000668corr1.
6. Elin RJ. Assessment of magnesium status for diagnosis and therapy. *Magn Res*. 2010;23:S194-8.
7. Ozarda Y. Reference intervals: current status, recent developments and future considerations. *Biochem Med (Zagreb)*. 2016;26:5-16.
8. Horowitz GL, Altaie S, Boyd JC, Ceriotti F, Garg U, Horn P, et al. Defining, establishing, and verifying reference intervals in the clinical laboratory. 3rd ed. CLSI guideline EP28-A3c. Clinical and Laboratory Standards Institute; 2010.
9. Workinger JL, Doyle RP, Bortz J. Challenges in the diagnosis of magnesium status. *Nutrients*. 2018;10:1202.
10. Colantonio DA, Kyriakopoulou L, Chan MK, Daly CH, Brinc D, Venner AA, et al. Closing the gaps in pediatric laboratory reference intervals: a CALIPER database of 40 biochemical markers in a healthy and multiethnic population of children. *Clin Chem*. 2012;58:854-68.
11. Tahmasebi H, Higgins V, Fung AWS, Truong D, White-Al Habeeb NMA, Adeli K. Pediatric reference intervals for biochemical markers: gaps and challenges, recent national initiatives and future perspectives. *EJIFCC*. 2017;28:43-63.
12. Ozarda Y, Higgins V, Adeli K. Verification of reference intervals in routine clinical laboratories: practical challenges and recommendations. *Clin Chem Lab Med*. 2018;57:30-7.
13. Roche Diagnostics GmbH. Magnesium Gen.2 (MG2): Method Sheet for Roche/Hitachi cobas c 701/702 Systems. Version 17.0. Mannheim, Germany: Roche Diagnostics GmbH; 2022.
14. Roche Diagnostics GmbH. PreciControl ClinChem Multi 1: Instructions for Use. Document No. 05117020001, Version 7.0. Mannheim, Germany: Roche Diagnostics GmbH.
15. Randox Laboratories Ltd. RIQAS General Clinical Chemistry External Quality Assessment Programme [Internet]. Crumlin, United Kingdom: Randox Laboratories Ltd. Available from: <https://www.randox.com/core-disciplines/external-quality-assessment/riqas-external-quality-assessment/clinical-chemistry-eqa>
16. Katayev A, Balciza C, Secombe DW. Establishing reference intervals for clinical laboratory test results: is there a better way? *Am J Clin Pathol*. 2010;133:180-6.
17. Ichihara K, Boyd JC; IFCC Committee on Reference Intervals and Decision Limits (C-RIDL). An appraisal of statistical procedures used in derivation of reference intervals. *Clin Chem Lab Med*. 2010;48:1537-51.
18. Linnet K. Nonparametric estimation of reference intervals by simple and bootstrap-based procedures. *Clin Chem*. 2000;46:867-9.
19. Harris EK, Boyd JC. On dividing reference data into subgroups to produce separate reference ranges. *Clinical Chemistry*. 1990;36:265-70.
20. Lahti A, Petersen PH, Boyd JC, Rustad P, Laake P, Solberg HE. Partitioning of nongaussian-distributed biochemical reference data into subgroups. *Clin Chem*. 2004;50:891-900.



Measurement of Irisin, Adiponectin, and Leptin Levels in Patients with Idiopathic Epilepsy

İdiopatik Epilepsili Hastalarda İrisin, Adiponektin ve Leptin Düzeylerinin Ölçülmesi

İb Şeyma SÖNMEZ ŞAHİN¹, İb Türkan UYGUR ŞAHİN², İb Tuğçe YILDIZ³, İb Zeynep Ebru ÇAKIN⁴, İb Seher DURAN⁵, İb Şahabettin SELEK⁶, İb Dilara Füsün İÇAĞASIOĞLU⁷, İb Akın İŞCAN⁸

¹Göztepe Prof. Dr. Süleyman Yalçın City Hospital, Clinic of Pediatric Neurology, İstanbul, Türkiye

²Altınbaş University Faculty of Medicine, Department of Pediatric Neurology, İstanbul, Türkiye

³Siirt Public Health Laboratory, Siirt, Türkiye

⁴University of Health Sciences Türkiye, Bağcılar Training and Research Hospital, Department of Pediatrics, İstanbul, Türkiye

⁵Bezmialem Vakıf University Faculty of Health Sciences, Department of Nutrition and Dietetics, İstanbul, Türkiye

⁶Bezmialem Vakıf University Faculty of Medicine, Department of Biochemistry, İstanbul, Türkiye

⁷Medipol University Faculty of Medicine, Department of Pediatric Neurology, İstanbul, Türkiye

⁸Private Practice, Pediatric Neurology, İstanbul, Türkiye

ABSTRACT

Objective: Epilepsy is a common childhood neurological disorder in which neuroinflammatory and metabolic mechanisms may contribute to epileptogenesis. Adipomyokines such as irisin, leptin, and adiponectin have been suggested to influence neuroinflammation, energy metabolism, and neuronal excitability. This study aimed to evaluate serum adipomyokine levels in children with idiopathic epilepsy and investigate their association with treatment response.

Methods: This case-control study included 68 children with idiopathic epilepsy (15 drug-resistant, 53 drug-responsive) and 66 healthy controls. Serum irisin, leptin, and adiponectin levels were measured using ELISA. After one-to-one matching based on body mass index (BMI), age, and sex, 58 epilepsy patients and 58 controls were included in the final analyses. For subgroup analyses, 15 drug-resistant patients were compared

Öz

Amaç: Epilepsi, çocukluk çağında sık görülen ve nöroenflamasyon ile metabolik süreçlerin epileptogenezde rol oynadığı düşünülen kronik bir nörolojik hastalıktır. İrisin, leptin ve adiponektin gibi adipomyokinlerin nöroenflamasyon, enerji metabolizması ve nöronal uyarılabilirlik üzerinde etkili olabileceği düşünülmektedir. Bu çalışmada idiyopatik epilepsili çocuklarda serum adipomyokin düzeylerinin değerlendirilmesi ve bu moleküllerin tedavi yanıtı ile ilişkilerinin araştırılması amaçlandı.

Yöntemler: Bu olgu-kontrol çalışmasına 68 idiyopatik epilepsili çocuk (15'i ilaç dirençli, 53'ü ilaç yanıtı) ve 66 sağlıklı kontrol dahil edildi. Serum irisin, leptin ve adiponektin düzeyleri ELISA yöntemiyle analiz edildi. Vücut kitle indeksi (VKİ), yaş ve cinsiyete göre birebir eşleştirme sonrası 58 epilepsi hastası ve 58 kontrol değerlendirmeye alındı. Alt grup analizlerinde 15 ilaç dirençli hasta ile VKİ uyumlu 15 ilaç yanıtı hasta karşılaştırıldı.

Address for Correspondence: Şeyma Sönmez Şahin, MD, Göztepe Prof. Dr. Süleyman Yalçın City Hospital, Clinic of Pediatric Neurology, İstanbul, Türkiye

E-mail: sonmezseyma@hotmail.com

ORCID IDs of the authors: Ş.S.Ş.: 0000-0002-2648-3683, T.U.Ş.: 0000-0002-7347-881X, T.Y.: 0000-0003-3885-9227, Z.E.Ç.: 0000-0002-9216-5878, S.D.: 0000-0001-6759-3566, Ş.S.: 0000-0003-1235-3957, D.F.İ.: 0000-0002-7728-2611, A.İ.: 0000-0001-5356-2646

Cite this article as: Sönmez Şahin Ş, Uygur Şahin T, Yıldız T, Çakin ZE, Duran S, Selek Ş, et al. Measurement of irisin, adiponectin, and leptin levels in patients with idiopathic epilepsy. Bezmialem Science. 2026;14(4):365-375

Received: 26.11.2025

Accepted: 06.07.2026

Published date: 26.10.2026



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

ABSTRACT

with 15 BMI-matched drug-responsive patients. Correlation, logistic regression, and receiver operating characteristic analyses were performed.

Results: After matching, serum irisin levels were significantly lower in epilepsy patients compared to controls ($p=0.032$). Leptin and adiponectin levels did not differ significantly between epilepsy and control groups. In subgroup analyses, leptin and adiponectin levels were significantly higher in drug-responsive patients than in drug-resistant patients ($p=0.043$ and $p=0.024$, respectively). Irisin levels did not differ between epilepsy subgroups. Logistic regression analyses showed that leptin and adiponectin were associated with treatment response.

Conclusion: The findings suggest that irisin may be involved in epilepsy-related mechanisms, whereas leptin and adiponectin may be associated with treatment response. Although adipomyokines appear to have limited diagnostic utility for differentiating epilepsy subtypes, they may represent potential biomarkers for predicting treatment response in pediatric epilepsy.

Keywords: Epilepsy, adipomyokines, irisin, leptin, adiponectin, drug resistance

ÖZ

Korelasyon, lojistik regresyon ve alıcı işletim karakteristiği analizleri yapıldı.

Bulgular: Eşleştirme sonrası epilepsi grubunda serum irisin düzeyleri kontrol grubuna göre anlamlı olarak daha düşük bulundu ($p=0,032$). Leptin ve adiponektin düzeyleri açısından epilepsi ve kontrol grupları arasında anlamlı fark saptanmadı. Alt grup analizlerinde leptin ve adiponektin düzeyleri ilaç yanıtı grupta ilaç dirençli gruba göre anlamlı olarak daha yüksekti (sırasıyla $p=0,043$ ve $p=0,024$). Irisin düzeyleri açısından alt gruplar arasında fark izlenmedi. Lojistik regresyon analizinde leptin ve adiponektin tedavi yanıtı ile ilişkili bulundu.

Sonuç: Bulgular, irisinin epilepsi ile ilişkili mekanizmalarda rol oynayabileceğini, leptin ve adiponektinin ise tedavi yanıtı ile ilişkili olabileceğini düşündürmektedir. Adipomyokinlerin epilepsi alt tiplerini ayırt etmede sınırlı tanısal değeri olmakla birlikte, çocukluk çağı epilepsisinde tedavi yanıtının öngörülmesinde potansiyel biyobelirteçler olabileceği düşünülmektedir.

Anahtar Kelimeler: Epilepsi, adipomyokin, irisin, leptin, adiponektin, ilaç direnci

Introduction

Epilepsy is a common neurological disorder characterized by recurrent unprovoked seizures and remains a significant cause of morbidity in children (1,2). According to the 2017 classification of the International League Against Epilepsy (ILAE), epilepsy arises from multiple etiologies, including structural, genetic, metabolic, infectious, and immune-related mechanisms (3). Increasing evidence suggests that neuroinflammation plays a key role in epileptogenesis, contributing to neuronal hyperexcitability and seizure persistence (4,5).

Adipose tissue and skeletal muscle function as endocrine organs that secrete bioactive molecules known as adipomyokines, which regulate energy metabolism, inflammation, and neuronal function (6,7). Among these, leptin, adiponectin, and irisin have attracted attention due to their potential roles in neuroprotection and modulation of neuronal excitability (8,9). Emerging evidence suggests that these molecules may also be involved in epileptogenesis, although their exact role remains unclear (6).

Despite growing interest, studies investigating adipomyokines in epilepsy are limited and show inconsistent findings, particularly in pediatric populations. Moreover, the relationship between adipomyokine levels and treatment response, especially drug-resistant epilepsy, has not been adequately explored.

Therefore, this study aimed to evaluate serum levels of irisin, leptin, and adiponectin in children with idiopathic epilepsy, to compare these levels between drug-responsive

and drug-resistant patients, and to investigate their association with current drug responsiveness.

Methods**Study Design and Participants**

This case-control study included 68 children aged 2-10 years with idiopathic epilepsy, recruited from the Pediatric Neurology Clinic of Bezmialem Vakıf University Hospital between October 2020 and March 2021. Fifteen children met the ILAE 2010 criteria for drug-resistant epilepsy (10), and 53 were classified as drug-responsive. The control group consisted of 66 healthy children without chronic diseases who attended the pediatric outpatient clinic for routine examination. All participants were in the prepubertal stage, as determined by Tanner criteria.

To minimize the potential confounding effect of body composition on adipomyokine levels, body mass index (BMI)-based matching was performed separately for the primary case-control analysis and for the subgroup analysis.

For the primary case-control analysis, patients with epilepsy and healthy controls were matched using a one-to-one optimal matching approach with a predefined BMI caliper of $\pm 1.5 \text{ kg/m}^2$. Matching was based only on BMI. Age and sex were not used as matching variables; however, they were compared after matching to evaluate the comparability of the groups. This procedure yielded 58 matched patients with epilepsy and 58 matched healthy controls (Figure 1).

For the subgroup analysis, a separate BMI-based matching procedure was performed to compare drug-resistant

and drug-responsive patients. Because this analysis did not include healthy controls, all available drug-resistant patients were considered as the reference group. Fifteen drug-responsive patients with eligible BMI values were selected for comparison, yielding 15 drug-resistant and 15 drug-responsive patients. Participants who could not be matched within the predefined BMI caliper were excluded only from the corresponding matched analysis (Figure 1).

Inclusion criteria were: prepubertal status, absence of structural brain lesions or genetic diagnoses, and no history of hypoxic/traumatic brain injury or chronic systemic illness. Exclusion criteria were: obesity (BMI $\geq$ 95th percentile for age and sex) (11), use of medications known to affect adipokine levels (other than anti-seizure drugs), recent acute illness, or abnormal brain magnetic resonance imaging. Patients with

a genetic diagnosis were also excluded. Only prepubertal patients were included because serum irisin and leptin levels are known to increase during puberty (12).

Epilepsy Classification

Epilepsy and seizure types were classified according to the ILAE 2017 criteria, which categorize seizures by onset (focal, generalized, unknown) and epilepsy as focal, generalized, combined focal and generalized, or unknown etiology (3). Drug responsiveness was determined according to the ILAE 2010 definition, where drug-resistant epilepsy is defined as the persistence of seizures despite appropriate use of at least two tolerated and appropriately chosen anti-seizure medications (10). Selection of anti-seizure drugs was based on patient age, seizure etiology, and electroencephalography (EEG) findings.

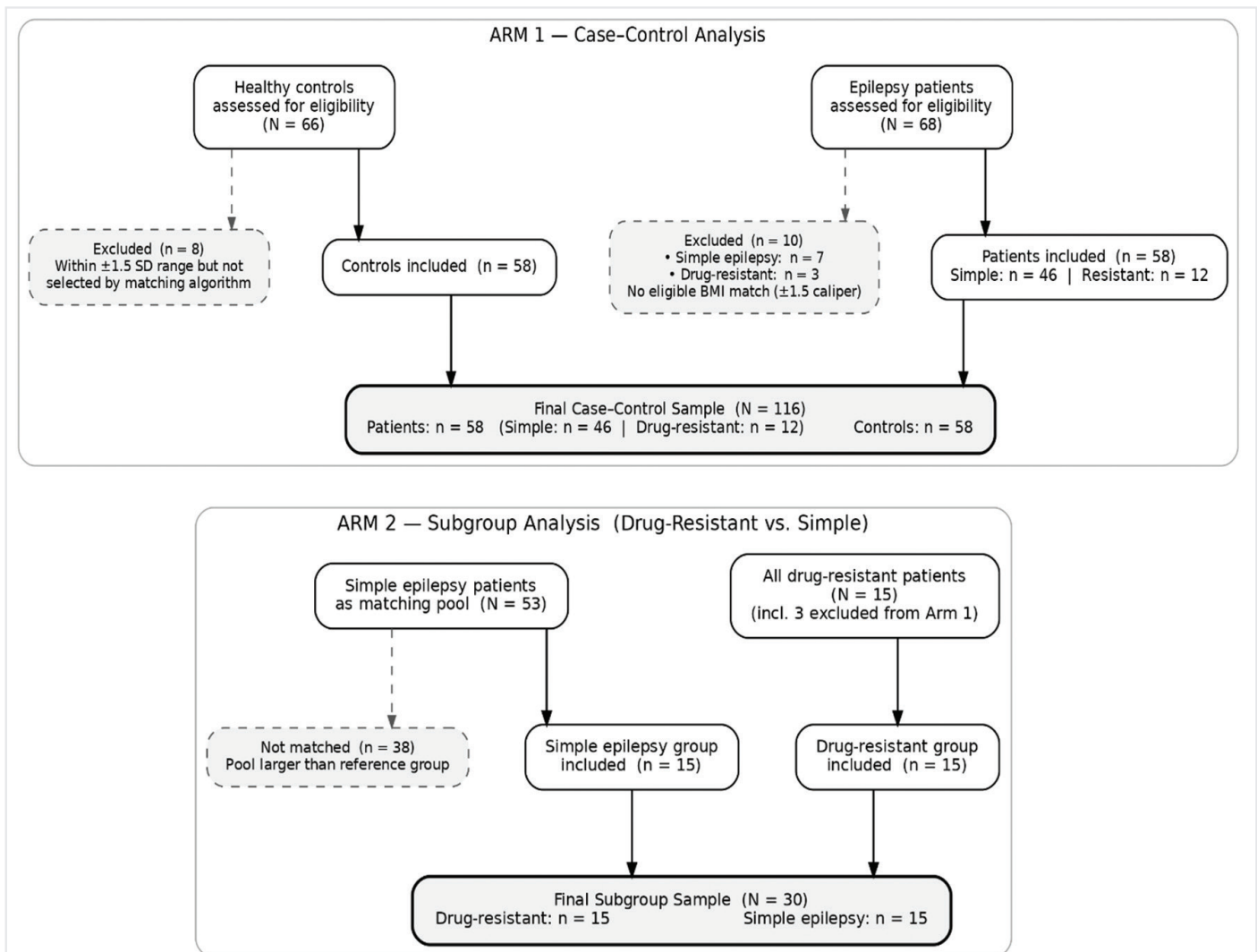


Figure 1. Flow diagram illustrating participant selection and BMI-matching procedures for the primary case-control analysis and the separate BMI-matched subgroup analysis. BMI matching was performed independently for each analysis. Consequently, three drug-resistant patients who were not included in the primary BMI-matched case-control analysis were included in the separate subgroup analysis, in which all 15 drug-resistant patients served as the reference group

BMI: Body mass index, SD: Standard deviation

Ethical Considerations

The study was approved by the Bezmialem Vakıf University Ethics Committee (decision no: 7/36, approval date: 22/07/2020, approval number: 8546). Written informed consent was obtained from the parents of all participants in accordance with the principles of the Declaration of Helsinki.

Clinical and Demographic Data

Patient histories were obtained from children and their caregivers, including sociodemographic characteristics, age at seizure onset, seizure types, disease duration, medication history, and seizure frequency. Physical and neurological examinations were performed by the same pediatric neurologist.

Anthropometric Measurements

BMI, arm circumference, and triceps skinfold thickness were measured by a single trained dietician. Arm circumference was measured at the midpoint between the acromion and olecranon processes using a non-stretchable tape. Triceps skinfold thickness was measured with a skinfold caliper over the triceps muscle.

Mid-arm circumference and triceps skinfold thickness measurements were available only for a subset of participants because these assessments were performed only when a trained dietitian was available during the recruitment period. Therefore, analyses involving these variables were based on the available data and should be interpreted with caution.

Blood Sampling Protocol

Venous blood samples were collected between 8:00-9:00 a.m., after an overnight fast of at least 8 hours, to minimize circadian and dietary influences. All participants refrained from strenuous physical activity for at least 24 hours before sampling. Where available, the time since the last seizure was recorded; if unavailable, this was noted as a study limitation. EEG was recorded during the interictal period using a 32-channel Nihon Kohden 1200 digital EEG device with the international 10–20 electrode placement system.

Biochemical Analysis

Blood was collected into plain gel separator vacutainer tubes and allowed to clot for 2 hours at room temperature (23 °C). Samples were centrifuged at 1000× g for 20 minutes, and the serum supernatant was aliquoted and stored at -80°C until analysis, strictly avoiding repeated freeze-thaw cycles. To minimize lot-to-lot variability, all samples were analyzed within a single batch. Serum irisin, leptin, and adiponectin concentrations were measured using commercially available ELISA kits (Human Irisin, Cat No: E-EL-H5735; Human LEP, Cat No: E-EL-H6017; Human ADP/Acrp30, Cat No: E-EL-H6122; all from Elabscience,

Houston, TX, USA) according to the manufacturer's protocols. Patient samples were assayed in singlicate.

The analytical characteristics of the kits were as follows: for irisin, the detection range was 78.13-5000 pg/mL with a sensitivity of 46.88 pg/mL; for leptin, the detection range was 15.63-1000 pg/mL with a sensitivity of 9.38 pg/mL; and for adiponectin, the detection range was 0.16-10 ng/mL with a sensitivity of 0.1 ng/mL. The intra-assay and inter-assay coefficients of variation for all three parameters were reported as <10%.

Statistical Analysis

All statistical analyses were performed using Jamovi (version 2.7.15; The jamovi project, 2025), which interfaces with R (Version 4.5; R Core Team, 2025). Normality of continuous variables was assessed using the Shapiro-Wilk test, and homogeneity of variances was evaluated with Levene's test.

For between-group comparisons, Student's independent samples t-test was applied when both normality and homogeneity of variance assumptions were satisfied; Welch's t-test was used when normality was met but variances were unequal; and the Mann-Whitney U test was employed when the normality assumption was violated. Categorical variables were compared using Pearson's chi-squared test or Fisher's exact test, the latter applied when expected cell frequencies fell below five.

Associations between continuous variables were examined using Spearman's rank correlation coefficient, given the non-normal distribution of several variables. To evaluate the association of irisin and adiponectin for group membership, separate univariable binary logistic regression models were fitted, with model fit evaluated using the likelihood ratio chi-squared statistic and McFadden's pseudo- R^2 .

Discriminative performance was quantified using the area under the receiver operating characteristic (ROC) curve (AUC). The statistical significance threshold was set at $\alpha=0.05$ for all analyses.

Due to the high proportion of censored leptin values (>1,000 pg/mL) in both groups, leptin was excluded from logistic regression and ROC analyses; it is reported descriptively only and findings should be considered exploratory.

Results

Participants

A total of 116 participants were included in the primary case-control analysis, comprising 58 patients with epilepsy and 58 healthy controls. The epilepsy group consisted of 46 patients with drug-responsive epilepsy and 12 with drug-resistant epilepsy after BMI-based matching.

Following one-to-one matching based on BMI, the patient and control groups were comparable with respect to age ($p=0.390$), sex distribution ($p=0.350$), and BMI ($p=0.754$), indicating successful matching (Table 1).

For the subgroup analysis, a separate BMI-matched cohort was constructed to compare drug-resistant and drug-responsive patients. This analysis included all 15 drug-resistant patients and 15 BMI-matched drug-responsive patients. The two subgroups did not differ significantly in age, sex, or BMI (all $p>0.05$), confirming the comparability of the matched subgroup (Table 2).

The participant selection and BMI-matching procedures for the primary case-control and subgroup analyses are illustrated in Figure 1.

Comparison Between Epilepsy and Control Groups

Among the available measurements, no statistically significant differences were observed in mid-arm circumference or triceps skinfold thickness between the epilepsy and control groups. However, these analyses were based on a limited number of participants and should therefore be interpreted with caution (Table 1).

Table 1. Comparison of demographic and biochemical variables between patient and control groups

Variable	Patient group (n=58)	Control group (n=58)	Test statistic	p-value
Sex, n (%)				
Male	23 (40%)	28 (48%)	$\chi^2=0.87$	0.350
Anthropometric measurements, mean $\pm$ SD ^a				
Mid-arm circumference (cm) (n=11/22)	15.93 $\pm$ 2.82	15.74 $\pm$ 1.92	t (31)=0.23	0.820
Triceps skinfold thickness (mm) (n=11/22)	10.30 $\pm$ 2.99	9.49 $\pm$ 2.17	t (31)=0.89	0.381
Continuous variables, median (IQR) ^b				
Age (years)	6.46 (4.4)	6.25 (2.9)	U=1526	0.390
BMI (kg/m ²)	16.15 (2.2)	16.00 (1.7)	U=1607	0.679
Irisin (pg/mL) ^c	14.90 (15.5)	27.83 (81.1)	U=515	0.032
Leptin (pg/mL) ^d	1000.00 (995.0)	89.89 (995.0)	U=1421	0.117
Adiponectin (μ g/mL)	2.12 (1.3)	1.17 (1.4)	U=1332	0.073

Bold p-values indicate statistical significance at $p<0.05$
 BMI: Body mass index, IQR: Interquartile range (Q3-Q1), SD: Standard deviation
^a: Mid-arm circumference and triceps skinfold thickness: normality confirmed by Shapiro-Wilk test ($p=0.160$ and 0.104 , respectively); Levene's test for equality of variances was non-significant. Student's t-test applied
^b: Normality assumption violated (Shapiro-Wilk $p<0.05$) for all remaining continuous variables; Mann-Whitney U test applied
^c: Levene's test also significant for irisin ($p=0.019$), indicating unequal variances in addition to non-normality. Mann-Whitney U test applied
^d: A substantial proportion of leptin values reached the upper assay limit (1000 pg/mL); therefore, leptin findings should be interpreted cautiously

Table 2. Comparison of demographic and biochemical variables between resistant and simple patient subgroups

Variable	Resistant group (n=15)	Simple group (n=15)	Test statistic	p-value
Sex, n (%)				
Male	7 (47%)	6 (40%)	$\chi^2=0.14$	0.710
Continuous variables, mean $\pm$ SD ^a				
Age (years)	6.79 $\pm$ 2.18	6.70 $\pm$ 2.55	t (28)=0.10	0.925
Adiponectin (μ g/mL)	1.61 $\pm$ 0.72	2.20 $\pm$ 0.60	t (27)=-2.40	0.024
Continuous variables, median (IQR) ^c				
BMI (kg/m ²)	17.10 (5.4)	17.20 (4.7)	U=111.5	0.983
Irisin (pg/mL)	16.27 (85.1)	23.30 (108.7)	U=52.0	1.000
Leptin (pg/mL)	224.40 (958.6)	1000.00 (0.0)	U=68.0	0.043

BMI: Body mass index, IQR: Interquartile range (Q3-Q1), SD: Standard deviation
^a: Normality confirmed (Shapiro-Wilk $p\geq 0.05$) and Levene's test non-significant: Student's t-test applied. Age: SW $p=0.065$; adiponectin: SW $p=0.110$.
^c: Normality assumption violated (Shapiro-Wilk $p<0.05$) for BMI, irisin, and leptin: Mann-Whitney U test applied
Note: Leptin IQR for the simple group is 0.0, reflecting a highly right-skewed distribution in which both Q1 and Q3 were at the upper detection limit (1,000.0 pg/mL). A substantial proportion of values in both groups were censored at this limit (57% of patients, 43% of controls), rendering statistical inference unreliable. All leptin findings are exploratory and must not be used for inferential conclusions

Among adipomyokines, irisin levels were significantly lower in the epilepsy group compared to controls (median: 14.90 vs. 27.83 pg/mL, $p=0.032$) (Table 1).

In contrast, no statistically significant differences were observed in leptin or adiponectin levels between the groups ($p=0.117$ and $p=0.073$, respectively). Because leptin measurements were affected by a substantial ceiling effect, this finding should be interpreted with caution (Table 1).

Comparison Between Drug-responsive and Drug-resistant Epilepsy

In the BMI-matched subgroup analysis, adiponectin levels were significantly higher in the drug-responsive group compared to the drug-resistant group (2.20 ± 0.60 vs. 1.61 ± 0.72 $\mu\text{g/mL}$, $p=0.024$) (Table 2).

Leptin values were also higher in the drug-responsive group (median: 1000.00 vs. 224.40 pg/mL). However, because a substantial proportion of leptin measurements reached the assay upper limit of quantification, these findings were considered exploratory and were not used for inferential conclusions (Table 2).

However, irisin levels did not differ significantly between the two subgroups ($p=1.000$) (Table 2).

Correlation Analyses

Spearman correlation analyses demonstrated a strong positive association between leptin and adiponectin levels ($\rho=0.84$, $p<0.001$).

In the full sample, mid-arm circumference was negatively correlated with both leptin ($\rho=-0.43$, $p=0.013$) and adiponectin ($\rho=-0.54$, $p=0.001$) levels, while BMI was positively associated with triceps skinfold thickness ($\rho=0.52$, $p=0.002$).

Within the epilepsy group, exploratory analyses showed that leptin level was positively associated with age at seizure onset ($\rho=0.31$, $p=0.019$) and negatively associated with treatment duration ($\rho=-0.35$, $p=0.008$). Adiponectin also showed a negative correlation with treatment duration ($\rho=-0.35$, $p=0.008$).

No significant correlations were observed between irisin levels and anthropometric or clinical variables.

Clinical Characteristics of Epilepsy Subgroups

There were no statistically significant differences between drug-responsive and drug-resistant groups in terms of clinical and demographic characteristics, including age at onset, gestational age, treatment duration, seizure type, EEG findings, or family history (all $p>0.05$) (Table 3).

Logistic Regression and ROC Analyses

In the case-control comparison, neither irisin nor adiponectin level was significantly associated with group membership in the univariable logistic regression analyses

(all $p>0.05$). Among these, irisin level showed the highest, although non-significant, discriminative performance (AUC: 0.643).

In the subgroup analysis, higher adiponectin levels were significantly associated with drug responsiveness (odds ratio: 3.67, $p=0.031$). Irisin level was not significantly associated with subgroup status. Because of the marked ceiling effect observed in leptin measurements, leptin was excluded from logistic regression and ROC analyses.

ROC curves for irisin and adiponectin in subgroup analyses are presented in Figure 2.

Discussion

The present study investigated the relationship between serum adipomyokine levels and treatment response in children with idiopathic epilepsy. The main findings were that irisin levels were significantly lower in patients with epilepsy than in healthy controls, whereas adiponectin levels were significantly higher in drug-responsive than in drug-resistant patients. Exploratory analyses also suggested higher leptin levels in the drug-responsive group; however, these findings should be interpreted with caution because of the marked ceiling effect observed in leptin measurements. These findings suggest that adipomyokines may be associated not only with epileptogenesis but also with differences in current drug responsiveness.

The pathophysiology of epilepsy involves multiple mechanisms, and increasing evidence highlights the role of inflammation in epileptogenesis. The inclusion of immune etiology in the ILAE classification and the demonstration of neuronal damage induced by pro-inflammatory mediators have further emphasized the importance of inflammation in epilepsy (3,4). This has increased interest in identifying immune-related molecules that may improve our understanding of disease mechanisms and potentially contribute to future research on disease monitoring and therapeutic strategies.

Recent reviews have further emphasized the interaction between metabolic pathways, neuroinflammation, and epileptogenesis, suggesting that adipokines may influence seizure susceptibility through oxidative stress, glucose metabolism, and blood-brain barrier regulation (6).

Leptin is a hormone primarily synthesized in adipocytes and is involved in energy balance, immune regulation, and neuronal function. It has been shown to influence hippocampal neuronal activity, synaptic plasticity, and neurotrophic processes (6,8,13,14). Previous studies have reported both proconvulsive and anticonvulsive effects of leptin, indicating that its effects may vary according to disease stage, receptor distribution, and metabolic conditions (15). Experimental studies have shown that leptin deficiency may increase seizure susceptibility, whereas leptin administration may exert neuroprotective effects (16,17).

Table 3. Clinical and demographic characteristics of resistant and simple epilepsy subgroups

Variable	Resistant group (n=15)	Simple group (n=15)	Test statistic	p-value
Continuous variables, median (IQR) ^a				
Age at onset (years)	4.00 (5.0)	5.50 (4.5)	U=81.0	0.197
Gestational age (weeks)	38.00 (0.0)	38.00 (0.0)	U=105.0	0.577
Treatment duration (years)	2.00 (3.1)	1.50 (1.5)	U=79.5	0.177
Categorical variables, n (%)				
Age at onset category			$\chi^2(2)=2.71$	0.269
Infant (1-24 months)	6 (40%)	5 (33%)		
Preschool (25-60 months)	5 (33%)	2 (13%)		
School-age (6-11 years)	4 (27%)	8 (53%)		
Gestational age				
Term	15 (100%)	13 (87%)	Fisher	0.483
Preterm	0 (0%)	2 (13%)		
Consanguineous marriage				
No	12 (80%)	14 (93%)	Fisher	0.598
Yes	3 (20%)	1 (6.7%)		
Mode of delivery				
NSD	13 (87%)	14 (93%)	Fisher	>0.999
C-section	2 (13%)	1 (6.7%)		
NICU admission history				
No	14 (93%)	14 (93%)	Fisher	>0.999
Yes	1 (6.7%)	1 (6.7%)		
Seizure type ^b				
Focal	5/13 (38%)	8/15 (53%)	$\chi^2(1)=0.62$	0.431
Generalised	8/13 (62%)	7/15 (47%)		
Epilepsy type ^c			$\chi^2(2)=3.66$	0.111
Focal	5/14 (36%)	8/11 (73%)		
Generalised	8/14 (57%)	3/11 (27%)		
Focal and generalised	1/14 (7.1%)	0/11 (0%)		
EEG finding ^d				
Focal	5/11 (45%)	7/10 (70%)	Fisher	0.387
Generalised	6/11 (55%)	3/10 (30%)		
History of febrile seizures				
No	12 (80%)	12 (80%)	Fisher	>0.999
Yes	3 (20%)	3 (20%)		
Family history of epilepsy				
No	10 (67%)	12 (80%)	$\chi^2(1)=0.68$	0.682
Yes	5 (33%)	3 (20%)		
Special education				
No	13 (87%)	14 (93%)	Fisher	>0.999
Yes	2 (13%)	1 (6.7%)		
Drug allergy	0 (0%)	0 (0%)	-	-
Genetic diagnosis	0 (0%)	0 (0%)	-	-
Additional disease				
No	15 (100%)	14 (93%)	Fisher	>0.999
Yes	0 (0%)	1 (6.7%)		

Table 3. Continued

Variable	Resistant group (n=15)	Simple group (n=15)	Test statistic	p-value
Autism features ^e				
No	14/15 (93%)	13/14 (93%)	Fisher	>0.999
Yes	1/15 (6.7%)	1/14 (7.1%)		
ADHD ^f				
No	13/14 (93%)	12/14 (86%)	Fisher	>0.999
Yes	1/14 (7.1%)	2/14 (14%)		

ADHD: Attention-deficit/hyperactivity disorder, IQR: Interquartile range (Q3-Q1), NICU: Neonatal intensive care unit, NSD: Normal spontaneous delivery, EEG:

Electroencephalography

^a: Mann-Whitney U test applied to all continuous variables; data are presented as median (IQR)

^b: Two patients in the resistant group had missing seizure-type data (analysed n=13/15)

^c: Five patients had missing epilepsy-type data (resistant n=14/15; simple n=11/15)

^d: Nine patients had missing EEG data (resistant n=11/15; simple n=10/15)

^e: One patient in the simple group had missing data (analysed n=14/15)

^f: Two patients had missing ADHD data (resistant n=14/15; simple n=14/15)

Note: No statistically significant between-group differences were detected for any variable (all p>0.05). Fisher's exact test was applied when expected cell frequencies were below 5; Pearson's chi-squared test was used otherwise

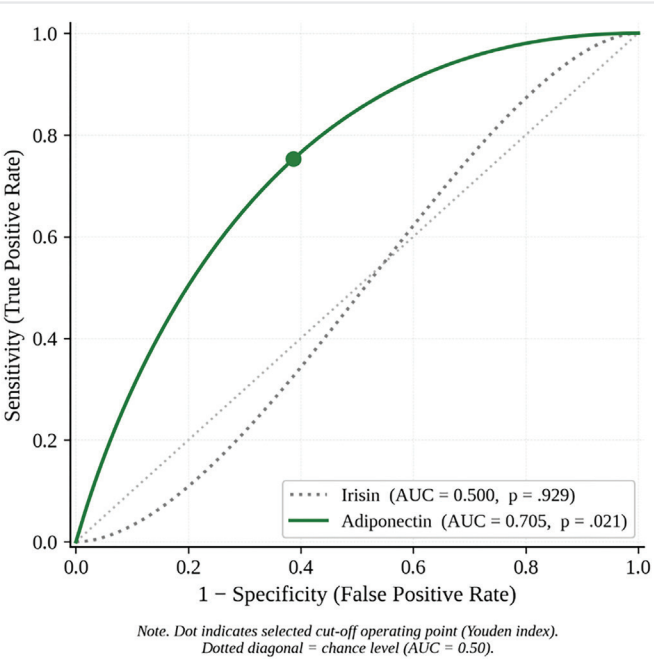


Figure 2. Receiver operating characteristic curves of irisin and adiponectin levels for discrimination between drug-responsive and drug-resistant epilepsy groups. Adiponectin demonstrated the highest discriminative performance among the evaluated adipomyokines

AUC: Area under the curve

In our study, leptin levels did not differ significantly between patients with epilepsy and healthy controls. Exploratory analyses suggested higher leptin levels in drug-responsive patients than in drug-resistant patients; however, these findings should be interpreted with caution because a substantial proportion of leptin measurements reached the upper limit of quantification of the assay. Therefore, no definitive conclusions regarding the association between leptin and current drug responsiveness can be drawn

from the present data. In addition, leptin levels showed a negative correlation with treatment duration, suggesting that circulating leptin concentrations may decrease during long-term treatment. Similar treatment-related alterations in leptin levels have been reported previously (18).

Adiponectin is another adipomyokine involved in energy homeostasis, neurogenesis, and synaptic plasticity. It has been shown to reduce oxidative stress, prevent neuronal injury, and preserve blood-brain barrier integrity after seizures (19,20). Several studies have demonstrated that adiponectin levels may be affected by antiepileptic drugs and ketogenic diet therapy (21-23). However, findings regarding its role in epilepsy remain inconsistent.

In the present study, adiponectin levels did not differ between patients with epilepsy and healthy controls but were significantly higher in drug-responsive than in drug-resistant patients. This finding suggests an association between higher adiponectin levels and current drug responsiveness in this cohort. In contrast to studies reporting lower adiponectin levels in drug-resistant epilepsy (24), our findings may reflect differences in study populations, treatment regimens, and methodological approaches.

Irisin is a recently identified adipomyokine with anti-inflammatory and neuroprotective properties. Previous studies have shown that irisin may contribute to neurogenesis and cognitive functions and may play a role in several neurological disorders (25,26). Several factors, including physical activity and metabolic status, are known to influence irisin levels (27).

In our study, irisin levels were significantly lower in epilepsy patients than in controls, suggesting a possible association between irisin and epilepsy-related mechanisms. However, no significant difference was observed between drug-responsive and drug-resistant groups. These findings are consistent with some previous studies reporting no

difference in irisin levels between epilepsy subgroups (15), although other studies reported elevated irisin levels in uncontrolled seizures (28). Differences in study populations, seizure types, and timing of blood sampling may explain these conflicting findings.

Emerging evidence suggests that irisin may act as a compensatory neuroprotective molecule during neuroinflammatory processes; however, its clinical significance in epilepsy has not yet been established (6).

In addition, irisin levels decreased with longer treatment duration in our study. This finding may be related to metabolic adaptations occurring during long-term treatment, although the underlying mechanisms remain unclear. Similar temporal changes have also been reported in experimental epilepsy models (29).

Correlation analyses revealed a strong positive association between leptin and adiponectin levels. However, because leptin measurements were affected by a substantial ceiling effect, this finding should be interpreted cautiously. Previous studies have also demonstrated interactions between adipomyokines, suggesting coordinated regulation of these molecules, although these relationships may vary depending on metabolic and disease-related conditions (30).

In the subgroup analysis, higher adiponectin levels were significantly associated with current drug responsiveness, whereas irisin showed no significant association. Because leptin measurements were substantially affected by the assay ceiling effect, leptin level was excluded from the logistic regression and ROC analyses, and no conclusions regarding its association with drug responsiveness can be drawn from these analyses. Overall, these findings suggest that adiponectin level may be associated with the current drug responsiveness status in this cohort, although confirmation in prospective studies is required. Consistent with previous reports, adipomyokines appear to have limited utility in distinguishing epilepsy subgroups (15,31).

Current evidence increasingly suggests that adipokines may be involved in mechanisms related to treatment response and disease progression rather than serving as diagnostic markers, particularly in pediatric epilepsy, where inflammatory and metabolic pathways may differ from those in adult (6).

To reduce the potential confounding effects of body composition on adipokine levels, BMI-matched analyses were performed in the present study. This approach may have contributed to a more reliable evaluation of the

relationship between adipomyokines and epilepsy-related variables.

Study Limitations

This study has several limitations. First, the relatively small sample size, particularly in the subgroup analyses, may have limited the statistical power to detect subtle between-group differences. Second, the lack of information regarding seizure activity within 24 hours before blood sampling may have influenced adipokine levels. Third, although medications affecting adipokine levels were excluded, factors such as physical activity and diet were not fully controlled. In addition, mid-arm circumference and triceps skinfold thickness measurements were available only for a subset of participants because these assessments could only be performed when a trained dietitian was available during the recruitment period. Therefore, findings related to these anthropometric variables should be interpreted with caution. Finally, because of the cross-sectional design, causal relationships cannot be established, and longitudinal studies are needed to clarify temporal changes in adipomyokines during the course of epilepsy and treatment. In addition, the marked ceiling effect observed for leptin measurements limited the interpretation of leptin-related findings, which should therefore be considered exploratory.

Conclusion

In conclusion, this study demonstrated that serum irisin levels were lower in children with idiopathic epilepsy than in healthy controls, whereas adiponectin levels were higher in drug-responsive than in drug-resistant patients. Exploratory analyses also suggested higher leptin levels in the drug-responsive group; however, these findings should be interpreted with caution because of the marked ceiling effect observed in leptin measurements. These findings suggest that adipomyokines may be associated not only with epilepsy-related inflammatory and metabolic mechanisms but also with differences in current drug responsiveness.

Although adipomyokines showed limited ability to distinguish between the study groups in this cohort, the observed association between adiponectin and current drug responsiveness warrants further investigation. Because of the cross-sectional design, these findings should not be interpreted as evidence of predictive performance. Prospective longitudinal studies with larger cohorts are needed to clarify the clinical significance of adipomyokines in pediatric epilepsy.

Ethics

Ethics Committee Approval: The study was approved by the Bezmialem Vakıf University Ethics Committee (decision no: 7/36, approval date: 22/07/2020, approval number: 8546).

Informed Consent: Written informed consent was obtained from the parents of all participants in accordance with the principles of the Declaration of Helsinki.

Footnotes

Authorship Contributions

Surgical and Medical Practices: Ş.S.Ş., T.U.Ş., Z.E.Ç., S.D., Ş.S., D.F.İ., Concept: Ş.S.Ş., T.U.Ş., T.Y., Ş.S., D.F.İ., A.İ., Design: Ş.S.Ş., T.U.Ş., T.Y., Ş.S., D.F.İ., A.İ., Data Collection or Processing: Ş.S.Ş., T.U.Ş., T.Y., Z.E.Ç., S.D., Ş.S., Analysis or Interpretation: Ş.S.Ş., T.U.Ş., T.Y., Z.E.Ç., Ş.S., D.F.İ., Literature Search: Ş.S.Ş., Writing: Ş.S.Ş., D.F.İ.

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

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

References

1. Kwan P, Brodie MJ. Refractory epilepsy: mechanisms and solutions. *Expert Rev Neurother*. 2006;6:397-406.
2. Hauser WA, Banerjee PN. Epidemiology of epilepsy in children. In: Pellock JM, Bourgeois BFD, Dodson WE, eds. *Pediatric Epilepsy: Diagnosis and Therapy*. 3rd ed. New York: Demos Medical Publishing. 2008. p.147-64.
3. Fisher RS, Cross JH, D'Souza C, French JA, Haut SR, Higurashi N, et al. Instruction manual for the ILAE 2017 operational classification of seizure types. *Epilepsia*. 2017;58:531-42.
4. Vezzani A, Fujinami RS, White HS, Preux PM, Blümcke I, Sander JW, et al. Infections, inflammation and epilepsy. *Acta Neuropathol*. 2016;131:211-34.
5. Shimada T, Takemiya T, Sugiura H, Yamagata K. Role of inflammatory mediators in the pathogenesis of epilepsy. *Mediators Inflamm*. 2014;2014:901902.
6. Shan Y, Chen Y, Gu H, Wang Y, Sun Y. Regulatory basis of adipokines leptin and adiponectin in epilepsy: from signaling pathways to glucose metabolism. *Neurochem Res*. 2023;48:2017-28.
7. Graf C, Ferrari N. Metabolic health-the role of adipomyokines. *Int J Mol Sci*. 2019;20:6159.
8. Mora-Muñoz L, Guerrero-Naranjo A, Rodríguez-Jiménez EA, Mastronardi CA, Velez-van-Meerbeke A. Leptin: role over central nervous system in epilepsy. *BMC Neurosci*. 2018;19:51.
9. Bloemer J, Pinky PD, Govindarajulu M, Hong H, Judd R, Amin RH, et al. Role of adiponectin in central nervous system disorders. *Neural Plast*. 2018;2018:4593530.
10. Kwan P, Arzimanoglou A, Berg AT, Brodie MJ, Allen Hauser W, Mathern G, et al. Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies. *Epilepsia*. 2010;51:1069-77.
11. Olcay N, Hülya G, Andrzej F. Bundak R, Gökçay G, Darendeliler F, ve ark. Türk çocuklarında vücut ağırlığı, boy uzunluğu, baş çevresi ve vücut kitle indeksi referans değerleri. *Çocuk Sağlığı ve Hastalıkları Dergisi*. 2008;51:1-14.
12. Kutlu E, Ozgen LT, Bulut H, Kocyigit A, Ustunova S, Hüseyinbas O, et al. Investigation of irisin's role in pubertal onset physiology in female rats. *Peptides*. 2023;163:170976.
13. Gorska-Ciebiada M, Ciebiada M. Association between serum irisin and leptin levels and risk of depressive symptoms in the diabetic elderly population. *J Clin Med*. 2023;12:4283.
14. Shan Y, Chen Y, Gu H, Wang Y, Sun Y. Regulatory basis of adipokines leptin and adiponectin in epilepsy: from signaling pathways to glucose metabolism. *Neurochemical Research*. 2023;48:2017-28.
15. Erkek OE, Milanlioğlu A, Komuroglu AU, Kara M, Huyut Z, Keskin S. Evaluation of serum ghrelin, nesfatin-1, irisin, and vasoactive intestinal peptide levels in temporal lobe epilepsy patients with and without drug resistance: a cross-sectional study. *Rev Assoc Med Bras (1992)*. 2021;67:207-12.
16. Erbayat-Altay E, Yamada KA, Wong M, Thio LL. Increased severity of pentylene tetrazol induced seizures in leptin deficient ob/ob mice. *Neurosci Lett*. 2008;433:82-6.
17. Guo Z, Jiang H, Xu X, Duan W, Mattson MP. Leptin-mediated cell survival signaling in hippocampal neurons mediated by JAK STAT3 and mitochondrial stabilization. *J Biol Chem*. 2008;283:1754-63.
18. Hamed SA, Fida NM, Hamed EA. States of serum leptin and insulin in children with epilepsy: risk predictors of weight gain. *Eur J Paediatr Neurol*. 2009;13:261-8.
19. Jeon BT, Shin HJ, Kim JB, Kim YK, Lee DH, Kim KH, et al. Adiponectin protects hippocampal neurons against kainic acid-induced excitotoxicity. *Brain Res Rev*. 2009;61:81-8.
20. Lee EB, Warmann G, Dhir R, Ahima RS. Metabolic dysfunction associated with adiponectin deficiency enhances kainic acid-induced seizure severity. *J Neurosci*. 2011;31:14361-6.
21. Aly RH, Amr NH, Saad WE, Megahed AA. Insulin resistance in patients on valproic acid: relation to adiponectin. *Acta Neurol Scand*. 2015;131:169-75.
22. Li HF, Zou Y, Xia ZZ, Gao F, Feng JH, Yang CW. Effects of topiramate on weight and metabolism in children with epilepsy. *Acta Paediatr*. 2009;98:1521-5.
23. Chyra M, Roczniak W, Świętochowska E, Dudzińska M, Oświęcimska J. The effect of the ketogenic diet on adiponectin, omentin and vaspin in children with drug-resistant epilepsy. *Nutrients*. 2022;14:479.
24. Toscano ECB, Lessa JMK, Gonçalves AP, Rocha NP, Giannetti AV, de Oliveira GN, et al. Circulating levels of adipokines are altered in patients with temporal lobe epilepsy. *Epilepsy Behav*. 2019;90:137-41.
25. Roca-Rivada A, Castela C, Senin LL, Landrove MO, Baltar J, Crujeiras AB, et al. FNDC5/irisin is not only a myokine but also an adipokine. *PLoS One*. 2013;8:e60563.
26. Pignataro P, Dicarlo M, Zerlotin R, Zecca C, Dell'Abate MT, Buccoliero C, et al. FNDC5/irisin system in neuroinflammation and neurodegenerative diseases: update and novel perspective. *Int J Mol Sci*. 2021;22:1605.

27. Flori L, Testai L, Calderone V. The "irisin system": from biological roles to pharmacological and nutraceutical perspectives. *Life Sci.* 2021;267:118954.
28. Elhady M, Youness ER, Gafar HS, Abdel Aziz A, Mostafa RSI. Circulating irisin and chemerin levels as predictors of seizure control in children with idiopathic epilepsy. *Neurol Sci.* 2018;39:1453-8.
29. Ergul Erkec O, Algul S, Kara M. Evaluation of ghrelin, nesfatin-1 and irisin levels of serum and brain after acute or chronic pentylentetrazole administrations in rats using sodium valproate. *Neurol Res.* 2018;40:923-9.
30. Jang HB, Kim HJ, Kang JH, Park SI, Park KH, Lee HJ. Association of circulating irisin levels with metabolic and metabolite profiles of Korean adolescents. *Metabolism.* 2017;73:100-8.
31. Özcan Ö, Arslan M, Güngör S, Yüksel T, Selimoğlu MA. Plasma leptin, adiponectin, neuropeptide Y levels in drug naive children with ADHD. *J Atten Disord.* 2018;22:896-900.



Comparative Analysis of Aesthetic Judgment between Orthodontists, General Dentists, and Different Versions of Large Language Models Using the IOTN-AC Index

IOTN-AC İndeksi Kullanılarak Ortodontistler, Genel Diş Hekimleri ve Büyük Dil Modellerinin Farklı Sürümleri Arasındaki Estetik Değerlendirmelerin Karşılaştırmalı Analizi

İD Rümeysa Cemre ÖNCÜ, İD Aylin PAŞAOĞLU BOZKURT

İstanbul Aydın University Faculty of Dentistry, Department of Orthodontics, İstanbul, Türkiye

ABSTRACT

Objective: This study aimed to compare the performance and diagnostic accuracy of general dentists and different artificial intelligence (AI)-based large language models (LLMs), including their older and newer versions, in photograph-based IOTN-AC index assessments, using orthodontists' evaluations as the reference standard.

Methods: The mean IOTN-AC scores derived from orthodontists were established as the reference (gold) standard. Cases were evaluated by general dentists and both earlier and newer versions of the ChatGPT, Claude, and Gemini families. Performance was analyzed using mean absolute error (MAE), Spearman's correlation, Lin's concordance correlation coefficient (CCC), systematic bias (Bland-Altman analysis), and diagnostic accuracy metrics (sensitivity/specificity) based on the clinical treatment need threshold of IOTN-AC ≥ 4.0 .

Results: General dentists demonstrated the closest agreement with the reference standard, exhibiting the lowest MAE (0.761) and the highest CCC (0.901, 95% confidence interval: 0.796-0.948). Conversely, all AI models exhibited a positive bias by systematically overestimating aesthetic impairment (+0.761 to +3.361). In the longitudinal analysis, although newer iterations of LLMs showed improved ranking performance (correlation), they regressed in absolute accuracy compared to earlier versions, exhibiting increased

ÖZ

Amaç: Bu çalışma, genel diş hekimlerinin ve yapay zeka (YZ) tabanlı farklı büyük dil modellerinin (LLM) eski ve yeni versiyonlarıyla birlikte fotoğraf tabanlı IOTN-AC indeksi değerlendirme performanslarını ve tanısal doğruluklarını ortodontistlerin referans standartlarına göre karşılaştırmayı amaçlamıştır.

Yöntemler: Ortodontistlerin ortalama IOTN-AC puanları referans (altın) standart olarak kabul edilmiştir. Olgular genel diş hekimleri ile ChatGPT, Claude ve Gemini ailelerinin eski ve yeni versiyonları tarafından değerlendirilmiştir. Performanslar; ortalama mutlak hata (MAE), Spearman korelasyonu, Lin'in konkordans korelasyon katsayısı (CCC), sistematik bias (Bland-Altman) ve IOTN-AC $\geq 4,0$ klinik eşiğine göre tanısal doğruluk (duyarlılık/özellik) metrikleri kullanılarak analiz edilmiştir.

Bulgular: Genel diş hekimleri, en düşük MAE (0,761) ve en yüksek CCC (0,901, %95 güven aralığı: 0,796-0,948) ile referans standarda en yakın uyumu göstermiştir. YZ modelleri ise estetik bozukluğu olduğundan yüksek değerlendirerek pozitif bias sergilemiştir (+0,761 ila +3,361). Boylamsal analizde, LLM'lerin yeni iterasyonları sıralama performansında (korelasyon) artış gösterse de, mutlak doğrulukta eski versiyonlara göre gerilemiş ve MAE değerleri artmıştır (örn. Gemini 3 Pro MAE 2,529'dan 3,406'ya yükselmiştir; $p=0,023$). IOTN-AC $\geq 4,0$ eşiğinde LLM'ler yüksek duyarlılık (%84,6-100,0) gösterirken,

Address for Correspondence: Rümeysa Cemre Öncü, İstanbul Aydın University Faculty of Dentistry, Department of Orthodontics, İstanbul, Türkiye

E-mail: rcemresahin@stu.aydin.edu.tr

ORCID IDs of the authors: R.C.Ö.: 0000-0002-7102-9324, A.P.B.: 0000-0001-8100-4684

Cite this article as: Öncü RC, Paşaoğlu Bozkurt A. Comparative analysis of aesthetic judgment between orthodontists, general dentists, and different versions of large language models using the IOTN-AC index. Bezmialem Science. 2026;14(4):376-384

Received: 06.06.2026

Accepted: 23.07.2026

Epub: 02.09.2026

Published date: 26.10.2026



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

ABSTRACT

MAE values (e.g., Gemini 3 Pro MAE increased from 2.529 to 3.406; $p=0.023$). At the IOTN-AC ≥ 4.0 threshold, while LLMs demonstrated high sensitivity (84.6-100.0%), they critically failed to distinguish healthy cases, resulting in extremely low specificity (e.g., 5.9% for Gemini 3 Pro).

Conclusion: Within the limitations of this study's sample and pooled evaluation design, general dentists demonstrated higher reliability and closer agreement with the reference standard than AI models in orthodontic aesthetic assessment. The architectural advancements in newer model versions did not translate positively into absolute clinical accuracy, creating a performance paradox. Although current LLMs show potential as preliminary screening tools due to their high sensitivity, they carry a significant risk of over-diagnosis owing to their low specificity, remaining far from being independent clinical decision support systems.

Keywords: Artificial intelligence, multimodal large language models, IOTN-AC, orthodontic treatment need, orthodontists, aesthetics

ÖZ

sağlıklı olguları ayırt etmede (özellik) kritik düzeyde başarısız olmuştur (örn., Gemini 3 Pro için %5,9).

Sonuç: Bu çalışmanın örnekleme ve havuzlanmış değerlendirme tasarımı sınırları içinde, genel diş hekimleri ortodontik estetik değerlendirmede YZ modellerinden daha yüksek güvenilirlik ve referans standarda daha yakın uyum göstermiştir. Modellerin yeni versiyonlarındaki mimari gelişim, mutlak klinik doğruluğa olumlu yansımamış ve bir performans paradoksu yaratmıştır. Güncel LLM'ler yüksek duyarlılıkları ile potansiyel bir tarama aracı olsalar da, düşük özellikleri nedeniyle ciddi bir "aşırı teşhis" (*over-diagnosis*) riski taşımakta ve henüz bağımsız bir klinik karar destek sistemi olmaktan uzak kalmaktadırlar.

Anahtar Kelimeler: Yapay zeka, çok modlu büyük dil modelleri, IOTN-AC, ortodontik tedavi ihtiyacı, ortodontistler, estetik

Introduction

Oral and dental health problems are among the most prevalent health conditions worldwide. Along with dental caries, periodontal diseases, and dental fluorosis, malocclusions represent one of the most frequently observed dental problems (1). Individuals with malocclusion may experience aesthetic concerns, speech difficulties, and masticatory dysfunction. Dental crowding is the primary reason patients seek orthodontic treatment (2). Achieving an attractive smile requires the correction of crowding and elimination of malocclusions, processes that are primarily managed by orthodontists (3,4).

Access to appropriate orthodontic treatment largely depends on the orthodontist's expertise and the quality of professional training. Orthodontic care may be pursued to improve both masticatory function and facial aesthetics, depending on the patient's awareness and expectations. General dentists play a crucial role in identifying patients who require referral, whereas orthodontists determine the necessity and scope of treatment. Postgraduate specialization has become increasingly important for enhancing clinical competence, improving diagnostic accuracy, promoting research awareness, and reducing the risk of complications (5).

Throughout history, dental aesthetics has held significant importance. Archaeological findings indicate that individuals have attempted to correct dental irregularities since early civilizations (6). Several indices and analytical tools have been developed to evaluate dentofacial aesthetics, including the Dental Aesthetic Index, Smile Index, Modified Smile Index, visual analog scale, and Q-sort analysis (7). However, the dental health component (DHC)

and aesthetic component (AC) of the Index of Orthodontic Treatment Need (IOTN) were specifically designed for orthodontic assessment (8).

In orthodontic practice, standardized indices with established reliability and validity are used to assess treatment need. The IOTN provides a structured numerical scoring system that quantifies the degree of deviation from ideal occlusion and is internationally accepted for evaluating orthodontic treatment requirements (9,10).

In recent years, rapid advancements in digital health technologies and artificial intelligence (AI) systems have significantly influenced the field of dentistry (11,12). AI systems support clinical workflows through image analysis, pattern recognition, and decision-support mechanisms (13). In orthodontics, three-dimensional modeling and predictive simulations enable the anticipation of tooth movements prior to treatment initiation, and automated systems contribute to aesthetic evaluation processes (14).

Large language models (LLMs), such as ChatGPT, Gemini, and Claude, were originally developed for natural language processing tasks. However, recent multimodal versions have incorporated visual processing capabilities in addition to text-based reasoning (13). Unlike conventional convolutional neural network-based image classification systems that are specifically trained for diagnostic imaging tasks, multimodal LLMs integrate visual encoders with large-scale reasoning architectures. Although previous AI studies in orthodontics have primarily relied on task-specific, trained image analysis models, the potential of widely accessible, general-purpose multimodal LLMs to perform structured aesthetic evaluations based on standardized clinical indices remains insufficiently explored.

For AI systems to be considered reliable clinical decision-support tools, it is essential to assess not only their agreement with human evaluators but also the consistency of their outputs across different model versions. Model updates may influence evaluation behavior, thereby affecting clinical predictability and stability.

Therefore, the present study aimed to evaluate and compare the levels of aesthetic discomfort perceived by orthodontists, general dentists, and multimodal LLM-based AI systems using the IOTN-AC index. Additionally, this study investigated inter-group consistency and examined intra-model stability by comparing the aesthetic scoring performance of earlier and updated versions of the selected LLM systems.

Methods

This study was conducted following approval from the Institutional Ethics Committee of İstanbul Aydın University (approval date: May 22, 2026, decision no: 193/2026). Given the novel nature of evaluating multimodal LLMs using the IOTN-AC index, formal a priori sample size calculation was not performed, and this research was designed as an exploratory study. Standardized frontal intraoral photographs obtained from 30 patients were included in this study. To minimize selection bias, the 30 patient cases were selected from the institutional database using a non-consecutive, simple random sampling method. The inclusion criteria strictly comprised: (1) high-quality and clear frontal intraoral photographs, (2) patients with no history of previous orthodontic treatment, and (3) the absence of any congenital or craniofacial syndromes. Conversely, cases presenting with poor image resolution, inadequate lighting, or significant framing errors were excluded from the study. Demographic data (e.g., age and gender) of the patients were not recorded, as the photographs were completely de-identified and the study focused exclusively on the objective morphological assessment of intraoral visual features. According to the reference standard evaluations, the IOTN-AC distribution of the selected cases was as follows: 19 mild (IOTN-AC 1-4), 7 moderate (IOTN-AC 5-7), and 4 severe (IOTN-AC 8-10) cases. The images were evaluated by 50 orthodontic specialists, 50 general dentists, and selected multimodal AI models. Each human evaluator assessed the same set of photographs in a single, independent session to determine the degree of aesthetic impairment. To evaluate the collective performance of the general dentist group, the pooling method was conducted by calculating the arithmetic mean of the 50 individual scores assigned to each specific photograph. This resulting mean value was then adopted as the representative final score of the general dentist group for that respective case.

Prior to evaluation, orthodontists and general dentists were provided with the internationally accepted IOTN-AC reference chart and were instructed to perform all assessments strictly according to this standardized scale.

The IOTN, developed by Brook and Shaw (9), is an objective method for assessing orthodontic treatment requirements. It evaluates malocclusion severity, degree, and aesthetic appearance (9,10). Following modifications recommended by the Swedish Medical Council, the index consists of two components: the DHC and the AC (15). The AC is based on a 10-point photographic scale representing varying levels of dental attractiveness. A score of 1 indicates little or no perceived treatment need, whereas a score of 10 reflects a definite need for orthodontic intervention (16,17). The DHC evaluates occlusal characteristics, emphasizing the most severe clinically significant trait when determining overall treatment need (9).

For AI evaluation, only multimodal LLM versions with integrated image interpretation capabilities were included. The AI evaluations were conducted following ethics committee approval, with all access and diagnostic assessments completed on May 25, 2026. The evaluations were performed using the premium web-based interfaces of the respective providers, without application programming interface integration. The specific models utilized were ChatGPT-4o and ChatGPT-5.2 (OpenAI; accessed via ChatGPT Plus subscription), Claude 3.5 Sonnet and Claude 4.5 Sonnet (Anthropic; accessed via Claude Pro subscription), and Gemini 1.5 Pro and Gemini 3 Pro (Google; accessed via Gemini Advanced subscription). Prior to assessment, the models were queried regarding their familiarity with the IOTN-AC index to ensure conceptual understanding. No task-specific retraining, fine-tuning, or dataset optimization was performed. This zero-shot evaluation approach was intentionally adopted to reflect real-world conditions of publicly accessible, general-purpose AI systems rather than specialized orthodontic image-analysis models. Both earlier and updated versions of the selected AI systems were evaluated independently. Each AI version assessed the identical photograph set only once in a single, independent run using the same IOTN-AC reference scale, without multiple iterations, prompt modifications, or regeneration of responses. Version-based comparisons were conducted to analyze the impact of model updates on aesthetic scoring behavior and to examine intra-model stability.

Orthodontists and general dentists were presented with standardized frontal intraoral photographs and were asked to identify the reference image on the IOTN-AC scale that most closely corresponded to each case. The same images were provided to the AI models, which were instructed to assign aesthetic scores according to the IOTN-AC index.

All selected photographs were high-resolution, clear, and standardized images. Data were organized in tabular format and presented as frequencies and percentages. Participation of clinicians was voluntary and based on informed consent.

Statistical Analysis

Statistical analyses were performed using Python-based data analysis, utilizing standard scientific computing libraries including Pandas, NumPy, SciPy, and Scikit-learn, with the significance level set at $p < 0.05$ for all tests. The normality of data distribution was assessed using the Shapiro-Wilk test. The mean IOTN-AC scores obtained from 50 orthodontists served as the reference standard. Agreement between the evaluator groups and the reference standard was analyzed using Spearman's rank correlation coefficient (ρ), Lin's concordance correlation coefficient (CCC), mean absolute error (MAE), and Bias (systematic deviation) metrics. Performance differences between AI model versions were compared using the Wilcoxon signed-rank test. Diagnostic performance metrics [sensitivity, specificity, positive predictive value (PPV), negative predictive value, and accuracy] were calculated based on the clinical treatment need threshold of IOTN-AC ≥ 4.0 . Limits of agreement between methods were evaluated using Bland-Altman analysis. Additionally, model performance was examined across case difficulty subgroups, stratified into "clear cases" [high consensus, standard deviation (SD) ≤ 0.508] and "ambiguous cases" (low consensus, SD > 0.508).

Results

The comparative performance metrics of general dentists and LLM-based AI systems, based on the mean IOTN-AC scores accepted as the reference standard from orthodontists, are presented in Table 1. According to the analysis results, general dentists exhibited the highest level of agreement with the orthodontists. This group

demonstrated the lowest (MAE: 0.761) and the highest Lin's (CCC: 0.901; 95% confidence interval: 0.796-0.948), indicating the absolute scoring closest to the reference standard in aesthetic assessment. In contrast, the absolute agreement of LLMs with orthodontist scores was found to be lower. Although the Spearman rank correlation coefficients (ρ) of all models were calculated to be statistically significant ($p < 0.05$), positive bias (systematic deviation) values indicating that aesthetic impairment was overestimated were detected in all models. Notably, in the Gemini 3 Pro model, this deviation exceeded +3 points, measured at a level that could potentially alter patients' IOTN treatment need classification.

The performance differences between older and newer versions of the same AI models were evaluated using the Wilcoxon signed-rank test (Table 2). The analysis revealed that in the newer versions of the ChatGPT and Gemini families (ChatGPT-5.2 and Gemini 3 Pro), Spearman correlation coefficients increased, indicating an improvement in ranking performance; however, in terms of absolute accuracy (MAE), error rates increased in the new iterations of all three model families. Specifically, Gemini 3 Pro showed a statistically significant increase in MAE compared to its predecessor, Gemini 1.5 Pro (MAE increased from 2.529 to 3.406; $p = 0.023$). The error rate also increased in the ChatGPT-5.2 model compared to ChatGPT-4o (+0.296 points), but this increase did not reach statistical significance ($p = 0.422$). The most pronounced performance loss between versions was detected in the Anthropic family. The newer model, Claude 4.5 Sonnet, exhibited a marked increase in error (+0.513 points; $p = 0.059$) compared to Claude 3.5 Sonnet. Furthermore,

Table 1. Agreement of general dentists and large language models with orthodontists' mean IOTN-AC scores

Rater group	Lin's CCC (95% CI)	Spearman's ρ (p-value)	MAE (95% CI)	Bias (95% CI)
General dentists (pooled)	0.901 (0.796 to 0.948)	0.949 (< 0.001)	0.761 (0.589 to 0.965)	+0.659 (0.437 to 0.898)
ChatGPT-4o	0.574 (0.243 to 0.786)	0.587 (< 0.001)	1.415 (0.929 to 2.043)	+0.761 (0.123 to 1.473)
ChatGPT-5.2	0.502 (0.292 to 0.677)	0.731 (< 0.001)	1.711 (1.309 to 2.159)	+1.494 (0.970 to 2.045)
Claude 3.5 Sonnet	0.528 (0.255 to 0.719)	0.584 (< 0.001)	1.861 (1.417 to 2.389)	+1.361 (0.733 to 2.019)
Claude 4.5 Sonnet	0.351 (0.095 to 0.565)	0.377 (0.040)	2.374 (1.856 to 2.932)	+1.394 (0.481 to 2.245)
Gemini 1.5 Pro	0.329 (0.127 to 0.519)	0.587 (< 0.001)	2.529 (2.001 to 3.190)	+2.361 (1.744 to 3.091)
Gemini 3 Pro	0.214 (0.079 to 0.382)	0.600 (< 0.001)	3.406 (2.752 to 4.114)	+3.361 (2.655 to 4.087)

Note: MAE indicates mean absolute error relative to orthodontists' mean scores. Bias represents the mean signed difference, with positive values indicating systematic overestimation of aesthetic severity. Spearman's ρ reflects rank-order agreement, while Lin's concordance correlation coefficient (CCC) evaluates both precision and absolute agreement.

IOTN: Index of Orthodontic Treatment Need, AC: Aesthetic component, MAE: Mean absolute error, CI: Confidence interval

Table 2. Comparison of model versions: analysis of performance change (Wilcoxon signed-rank test)

Model comparison (old → new)	MAE (old)	MAE (new)	Change in MAE	Result	p-value*
ChatGPT-4o → ChatGPT-5.2	1.415	1.711	+0.296	Degradation	0.422
Claude 3.5 Sonnet → Claude 4.5 Sonnet	1.861	2.374	+0.513	Degradation	0.059
Gemini 1.5 Pro → Gemini 3 Pro	2.529	3.406	+0.877	Significant degradation	0.023

Note: MAE: Mean absolute error (lower is better). Change in MAE: Positive values indicate an increase in error (worsening performance)

*: Statistically significant difference ($p < 0.05$)

Claude 4.5 Sonnet demonstrated a weaker agreement with the reference standard, showing a higher MAE (2.37 vs. 1.86), a lower Spearman correlation coefficient (0.38 vs. 0.58), and a decreased Lin's CCC (0.35 vs. 0.53) with wider confidence intervals.

Bland-Altman analysis demonstrated differences in individual reliability among the groups (Figure 1). General

dentists exhibited high individual agreement with the reference standard, showing low systematic error (Bias: +0.659) and narrow limits of agreement (-0.64 to +1.96). In all AI models, a systematic tendency toward "overestimation" (positive bias) was detected, and this deviation reached a level of +3.36 in the Gemini 3 Pro model. Additionally, the limits of agreement for AI models were found to be wider compared to human evaluators. For example, the wide

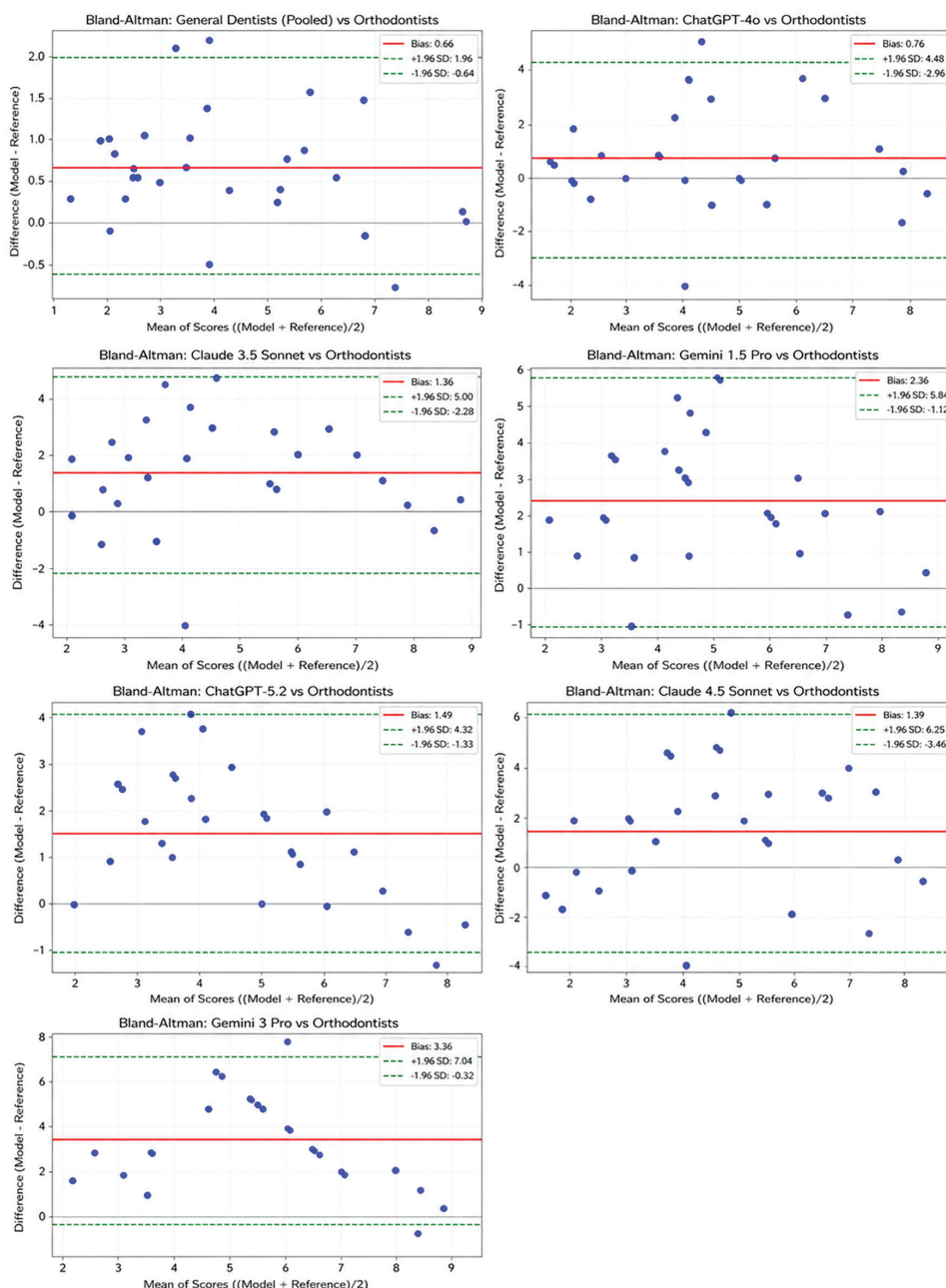


Figure 1. Bland-Altman plots evaluating the agreement between the raters (general dentists and AI models) and the reference standard (orthodontists). The solid red line represents the mean bias (systematic difference between the rater and the reference standard). The upper and lower dashed green lines indicate the 95% limits of agreement, calculated as mean bias ± 1.96 SDs. The x-axis represents the mean of the scores given by the specific rater and the reference standard $[(\text{model} + \text{reference})/2]$, while the y-axis shows the absolute difference between them (model - reference)]

SD: Standard deviation, AI: Artificial intelligence

range calculated for ChatGPT-4o (-2.95 to +4.47) indicates a fluctuation of approximately 7 points in individual cases. In the graphical distributions in Figure 1, proportional bias was observed, where the magnitude of error systematically changed with case severity.

The results of the bias analysis stratified by case severity indicated that the error characteristics of the AI models varied according to case categories (Table 3). While general dentists exhibited a positive bias of +0.78 and +0.71 in mild (IOTN 1-4) and moderate (IOTN 5-7) severity cases respectively, this value was calculated as -0.21 in severe cases (IOTN 8-10). In the AI models, an inverse relationship was detected between case severity and error magnitude. All models showed a distinct "overestimation" tendency in mild cases; this deviation reached +4.16 points in the Gemini 3 Pro model and +2.95 points in the Gemini 1.5 Pro model. In severe cases, the bias values of the same models decreased significantly, approaching zero or turning negative (e.g., Claude 3.5 Sonnet: -0.01; ChatGPT-4o: -0.67).

Model performances were evaluated under varying levels of clinical consensus (expert agreement), and the dataset was stratified into "clear cases" (high expert consensus, SD ≤ 0.508 , n=15) and "ambiguous cases" (low expert consensus,

SD > 0.508 , n=15) (Table 4). According to the analysis results, general dentists exhibited low error rates in both categories; the MAE was 0.976 in clear cases and 0.547 in ambiguous cases. In the AI models, while higher error rates and positive bias (e.g., Gemini 3 Pro: +3.911; ChatGPT-5.2: +2.244) were detected in clear cases where experts had a strong consensus, error margins were observed to decrease in ambiguous cases where expert consensus was lower. For instance, while ChatGPT-4o had a bias of +1.311 in the clear cases group, this value was measured at +0.211 in the ambiguous cases group, which was lower than the bias value of general dentists in the same group (+0.355). This trend was consistent across other AI models, and MAE values were systematically calculated to be lower in the ambiguous cases category (e.g., ChatGPT-5.2 MAE: 2.263 in clear cases vs. 1.160 in ambiguous cases).

An IOTN-AC ≥ 4.0 threshold was utilized to determine treatment need in the diagnostic performance evaluation (Table 5). The analysis revealed that general dentists demonstrated 80.0% accuracy and 70.6% specificity. The sensitivity rates of the AI models were measured between 84.6% and 100%, with the ChatGPT-5.2 and Gemini 3 Pro models achieving 100% sensitivity. Conversely, the specificity values of the models were found to be low; this

Table 3. Systematic bias stratified by case severity (IOTN-AC categories)

Rater group	Mild cases ^a	Moderate cases ^b	Severe cases ^c
General dentists (pooled)	+0.780	+0.710	-0.210
ChatGPT-4o	+1.370	-0.160	-0.670
Claude 3.5 Sonnet	+1.740	+0.970	-0.010
ChatGPT-5.2	+2.210	+0.720	-1.010
Claude 4.5 Sonnet	+1.900	+1.090	-1.010
Gemini 1.5 Pro	+2.950	+1.970	-0.340
Gemini 3 Pro	+4.160	+2.590	+0.330

Note: Values represent mean bias, calculated as the rater score minus the orthodontist score, where positive values indicate overestimation of aesthetic severity and negative values indicate underestimation
^a: Mild cases: reference standard mean score < 4.5 (n=19), ^b: Moderate cases: reference standard mean score ≥ 4.5 and ≤ 7.5 (n=7), ^c: Severe cases: reference standard mean score > 7.5 (n=4), IOTN: Index of Orthodontic Treatment Need, AC: Aesthetic component

Table 4. Comparison of performance metrics in "clear cases" (low observer variability) versus "ambiguous cases" (high observer variability)

Rater group	Clear cases (low SD) MAE	Clear cases bias	Ambiguous cases (high SD) MAE	Ambiguous cases bias
General dentists	0.976	+0.963	0.547	+0.355
ChatGPT-4o	1.567	+1.311	1.264	+0.211
ChatGPT-5.2	2.263	+2.244	1.160	+0.744
Claude 3.5 Sonnet	1.897	+1.577	1.824	+1.144
Claude 4.5 Sonnet	2.260	+1.844	2.488	+0.944
Gemini 1.5 Pro	2.855	+2.711	2.203	+2.011
Gemini 3 Pro	3.911	+3.911	2.901	+2.811

Note: Clear cases (high consensus) are defined as images where the standard deviation of orthodontist scores was less than or equal to 0.508 (n=15), indicating strong agreement among experts. Ambiguous cases (low consensus) are defined as images where the standard deviation exceeded 0.508 (n=15), indicating disagreement or clinical ambiguity
MAE: Mean absolute error (lower is better), Bias: Mean difference (positive values indicate overestimation), SD: Standard deviation

rate was calculated as 5.9% for Gemini 3 Pro and 11.8% for Gemini 1.5 Pro. The PPVs of the AI models ranged from 44.4% to 63.2%.

Discussion

This study is among the first investigations directly compares orthodontists, general dentists, and LLM-based AI systems in aesthetic assessment based on the IOTN-AC index. While traditional orthodontic diagnosis relies heavily on clinical experience and professional judgment, AI is increasingly being integrated into diagnosis and treatment processes by analyzing large datasets. However, the IOTN-AC index is not solely based on measurable morphological parameters; it relies on a subjective assessment shaped by aesthetic sensibility and the interpretation of visual stimuli (18). By exhibiting the highest agreement with orthodontists, general dentists have supported the central role of clinical training and experience-based judgment in aesthetic assessment. This result suggests that aesthetic perception is likely not merely dependent on visual pattern recognition, but on an integrated clinical understanding that remains uniquely human. In this context, the role of LLMs in such clinical tasks that require human perception rather than objective measurement remains a subject of significant debate (18,19).

In our study, LLMs were conceptualized not as technical tools intentionally designed to classify images objectively or simply as “seeing” systems, but as multimodal systems that shape human aesthetic perception, interpreting, describing, and categorizing visual information in a human-like manner. From this perspective, the primary objective of the study was not to test image recognition accuracy, but to explore whether LLMs could meaningfully represent aesthetic severity within a human-centered conceptual framework (19).

The wide limits of agreement and distinct proportional error observed in LLMs suggest significant limitations regarding the clinical reliability of these models. In particular, the marked increase in the margin of error as case severity decreases suggests that the models experience serious

difficulties in distinguishing healthy cases with low visual complexity in the present dataset. This pattern indicates that although LLMs can approximate certain aspects of human perceptual reasoning, they lack the calibration required for reliable absolute clinical scoring. These deviations, which can exceed several points in some cases, have the potential to significantly alter IOTN-based clinical classification. Therefore, it is not sufficient for current LLMs to merely demonstrate average accuracy; the high variability at the individual case level limits the use of these systems as independent clinical decision-makers at the current stage.

The most striking finding of our study, which contradicts general expectations in the literature, is the systematic regression observed in IOTN index scoring of new generation LLMs compared to their predecessors. In particular, the statistically significant performance loss observed in the Gemini 3 Pro model (p=0.023), along with similar deterioration trends in other models, suggests that advances in AI development do not inherently and always equate to “clinical improvement”. This finding shows that increased model complexity does not guarantee enhanced performance in perception-based medical tasks and serves as a critical warning for medical AI integration. It is becoming a necessity that LLM-based tools designed for healthcare systems undergo *de novo* validation tests after every software update, just like the validation of an entirely new medical device. This paradoxical result can be attributed to the “reinforcement learning from human feedback” and “safety alignment” strategies, which are increasingly prioritized in the training processes of current LLMs (20,21).

Another fundamental weakness of LLMs in our study was their low specificity. The fact that the models demonstrated a sensitivity approaching that of experts in detecting severe malocclusions (IOTN 8-10) suggests that AI has the capacity to recognize distinct anatomical deviations such as severe crowding or overjet. Although sensitivity rates reaching 100% indicate that the models do not miss cases requiring treatment, this success is largely overshadowed by specificity rates dropping as low as 5.9%. This situation may

Table 5. Diagnostic accuracy metrics for orthodontic treatment need (threshold: IOTN-AC ≥4.0)					
Rater group	Sensitivity	Specificity	PPV	NPV	Accuracy
General dentists	0.923	0.706	0.706	0.923	0.800
ChatGPT-4o	0.923	0.588	0.632	0.909	0.733
ChatGPT-5.2	1.000	0.176	0.481	1.000	0.533
Claude 3.5 Sonnet	0.846	0.412	0.524	0.778	0.600
Claude 4.5 Sonnet	0.923	0.412	0.545	0.875	0.633
Gemini 1.5 Pro	0.923	0.118	0.444	0.667	0.467
Gemini 3 Pro	1.000	0.059	0.448	1.000	0.467
Note: Threshold for treatment need is defined as an IOTN-AC score of 4.0 or higher Sensitivity: Ability to identify cases requiring treatment, Specificity: Ability to identify healthy cases correctly, PPV: Positive predictive value, NPV: Negative predictive value, IOTN: Index of Orthodontic Treatment Need, AC: Aesthetic component					

indicate that AI acts more like an overly sensitive and flawed alarm system rather than an objective diagnostic tool. The systematic and over-scoring tendency observed in mild cases (IOTN 1-4) indicates that these systems currently lack the calibration needed to distinguish the fine line between “clinical norm” and “pathology”. This tendency towards “over-diagnosis” can lead to a significant burden resulting from unnecessary specialist referrals in clinical practice. Considering the superior level of specificity exhibited by general dentists, LLMs should be utilized as auxiliary tools under professional supervision solely to benefit from their high sensitivity. While the models erroneously labeling mild cases as “severe” creates an unnecessary referral burden on the system and unwarranted anxiety for the patient, the success of general dentists suggests that human common sense and the principle of *primum non nocere* (first, do no harm) cannot yet be replicated by algorithms. The safety-oriented approach of the developers appears to have created a strict error-avoidance behavior in the models; this may have replaced technical accuracy with an “overprotective” and artificial score inflation that perceives even the slightest deviations as a risk.

While LLMs failed in low-variation “clear cases” where orthodontists reached full consensus, they exhibited a performance approaching, and in some cases even exceeding, that of general dentists in high-variation “ambiguous cases” characterized by expert disagreements. The natural indecision among human observers in “ambiguous cases”, which generally represent individuals with borderline treatment needs, can cause AI predictions to be recorded statistically as “different but valid opinions” rather than “errors”. Furthermore, in the face of the density of “ideal aesthetic” and “severe anomaly” images in LLM training datasets, the more limited data on “borderline” cases may have led the models to manage this ambiguity through an averaging method known as the “regression effect”; this situation suggests that AI’s strategy for managing clinical uncertainty is fundamentally different from that of humans, exhibiting a “generalizing” or averaging behavior where humans adopt a “conservative” approach in the face of uncertainty.

When compared with current studies in the literature, our findings add a deeper and more critical dimension to the ongoing debates regarding the role of AI in orthodontic diagnosis. While Kazimierczak et al. (22) paint a general picture of optimism by highlighting the potential of AI to revolutionize orthodontic diagnosis and treatment planning, our study reveals with concrete data that this “revolution” is not yet complete, especially when it comes to subjective aesthetic perception.

In particular, when examining the relationship between model performance and version updates, our findings form a striking contrast with the results of Kiliç and Mansız (23). While Kiliç and Mansız (23) reported that the reliability and readability of orthodontic information

increased as text-based ChatGPT versions were updated, our study found that new generation models focusing on visual analysis capabilities (e.g., Gemini 3 Pro) experienced a statistically significant performance loss compared to their predecessors. This situation suggests that text-based improvements do not directly translate to visual aesthetic judgment; on the contrary, factors such as “safety alignment” can lead to a paradoxical regression in visual analysis.

Similarly, Perez-Pino et al. (24) stated that virtual assistants exhibited inconsistencies in routine questions and required professional validation; our study supports this view by indicating that a potential reason for this inconsistency is the extreme sensitivity and systematic “over-diagnosis” tendency developed by the models due to the strict safety constraints.

In addition, Stetzel’s (25) thesis study on IOTN-AC estimation pointed out the difficulties in automating this index. Our study, on the other hand, suggested that this difficulty might stem not only from a technical inadequacy but also from the models’ potential inability to fully imitate the “contextual flexibility” specific to human perception (as we see in general dentists).

Study Limitations

This study has certain limitations. First, given the novel and exploratory nature of evaluating multimodal LLMs using the IOTN-AC index, a formal a priori sample size calculation (power analysis) was not performed. Consequently, the relatively small sample size limits the broad generalizability of the findings, and the present research should be considered primarily as an exploratory or pilot investigation. Second, aesthetic assessments were conducted solely on standardized two-dimensional frontal intraoral photographs obtained from the 30 included patients, and the DHC of the IOTN index, which evaluates overall treatment need, was not included in the analysis. Third, when testing the AI models, no task-specific retraining, fine-tuning, or dataset optimization was performed; instead, a “zero-shot” evaluation approach was intentionally adopted to reflect the real-world conditions of publicly accessible systems. While this methodological choice reflects the general-purpose capabilities of the models, it does not represent the potential performance of algorithms specifically optimized for orthodontic image analysis.

Conclusion

In conclusion, within the specific limitations of this study’s sample and pooled evaluation design although current LLM systems show promise in identifying pathology (sensitivity), their weakness in distinguishing healthy cases (specificity) suggests that they may still be inadequate at fully replicating the depth and contextual sensitivity of human aesthetic judgment. This situation limits the use of these models as autonomous screening tools; rather, it supports

their positioning as preliminary assessment adjuncts that require clinician validation. Human expertise remains the ultimate reference standard in aesthetic orthodontic assessment; therefore, the integration of LLMs into clinical workflows should be approached with caution, carefully considering these inherent structural limitations.

Ethics

Ethics Committee Approval: This study was conducted following approval from the Institutional Ethics Committee of İstanbul Aydın University (approval date: May 22, 2026, decision no: 193/2026).

Informed Consent: Participation of clinicians was voluntary and based on informed consent.

Footnotes

Authorship Contributions

Surgical and Medical Practices: R.C.Ö., A.P.B., Concept: R.C.Ö., A.P.B., Design: R.C.Ö., A.P.B., Data Collection or Processing: R.C.Ö., Analysis or Interpretation: R.C.Ö., A.P.B., Literature Search: R.C.Ö., Writing: R.C.Ö., A.P.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. Sureshbabu A, Chandu G, Shafiulla M. Prevalence of malocclusion and orthodontic treatment needs among 13-15-year-old schoolchildren of Davangere city, Karnataka. *J Indian Assoc Public Health Dent.* 2005;5:32.
2. Buschang PH, Shulman JD. Incisor crowding in untreated persons 15-50 years of age: United States, 1988-1994. *Angle Orthod.* 2003;73:502-8.
3. Sam G. Orthodontics as a prospective career choice among undergraduate dental students: a prospective study. *J Int Soc Prev Community Dent.* 2015;5:290-5.
4. Dewey M. Who is an "orthodontist"? 1915. *Am J Orthod Dentofacial Orthop.* 2015;147:523-4.
5. Paquette JM, Sheets CG. The second "DDS" degree: a formula for practice success. *J Am Dent Assoc.* 2004;135:1321-5.
6. Wahl N. Orthodontics in 3 millennia. Chapter 1: antiquity to the mid-19th century. *Am J Orthod Dentofacial Orthop.* 2005;127:255-9.
7. American Dental Association, National Commission on Recognition of Dental Specialties and Certifying Boards. Recognized dental specialties. American Dental Association. Published 2018. [cited September 2, 2026], 2025. <https://www.ada.org>
8. Evans R, Shaw W. Preliminary evaluation of an illustrated scale for rating dental attractiveness. *Eur J Orthod.* 1987;9:314-8.
9. Brook PH, Shaw WC. The development of an Index of Orthodontic Treatment priority. *Eur J Orthod.* 1989;11:309-20.
10. Shaw WC, Richmond S, O'Brien KD. The use of occlusal indices: a European perspective. *Am J Orthod Dentofacial Orthop.* 1995;107:1-10.
11. 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.
12. Sener E, Baksi Sen G. Role of artificial intelligence applications in dentomaxillofacial radiology: part 1 [Dentomaksillofasiyal radyolojide yapay zeka uygulamalarının rolü: bölüm 1]. *Selcuk Dent J.* 2022;9:713-20. Turkish.
13. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. *Nat Med.* 2023;29:1930-40.
14. Richmond S, Shaw WC, O'Brien KD, Buchanan IB, Jones R, Stephens CD, et al. The development of the PAR Index (Peer Assessment Rating): reliability and validity. *Eur J Orthod.* 1992;14:125-39.
15. Hatia A, Doldo T, Parrini S, Chisci E, Cipriani L, Montagna L, et al. Accuracy and completeness of ChatGPT-generated information on interceptive orthodontics: a multicenter collaborative study. *J Clin Med.* 2024;13:735.
16. Richmond S, Aylott NA, Panahei ME, Rolfe B, Tausche E. A two-center comparison of orthodontists' perceptions of orthodontic treatment difficulty. *Angle Orthod.* 2001;71:404-10.
17. de Almeida AB, Leite IC. Orthodontic treatment need for Brazilian schoolchildren: a study using the Dental Aesthetic Index. *Dental Press J Orthod.* 2013;18:103-9.
18. Wang L, Li J, Zhuang B, Huang S, Fang M, Wang C, et al. Accuracy of large language models when answering clinical research questions: systematic review and network meta-analysis. *J Med Internet Res.* 2025;27:e64486.
19. Makrygiannakis MA, Giannakopoulos K, Kaklamanos EG. Evidence-based potential of generative artificial intelligence large language models in orthodontics: a comparative study of ChatGPT, Google Bard, and Microsoft Bing. *Eur J Orthod.* 2025;48:cjae017.
20. Casper S, Davies X, Shi C, Gilbert TK, Scheurer J, Rando J, et al. Open problems and fundamental limitations of reinforcement learning from human feedback. Preprint. arXiv. 2023.
21. Ouyang L, Wu J, Jiang X, Almeida D, Wainwright CL, Mishkin P, et al. Training language models to follow instructions with human feedback. Preprint. arXiv. 2022.
22. Kazimierczak N, Kazimierczak W, Serafin Z, Nowicki P, Nożewski J, Janiszewska-Olszowska J. AI in orthodontics: revolutionizing diagnostics and treatment planning-a comprehensive review. *J Clin Med.* 2024;13:344.
23. 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.
24. Perez-Pino A, Yadav S, Upadhyay M, Cardarelli L, Tadinada A. The accuracy of artificial intelligence-based virtual assistants in responding to routinely asked questions about orthodontics. *Angle Orthod.* 2023;93:427-32.
25. Stetzel L. Artificial intelligence (AI) for predicting the aesthetic component (AC) of the Index of Orthodontic Treatment Need (IOTN) (dissertation). Columbus, Ohio: The Ohio State University. 2023.



AHEAD Score and Prolonged Intensive Care Unit Stay Following Isolated Coronary Artery Bypass Grafting

İzole Koroner Arter Bypass Greftleme Sonrası AHEAD Skoru ile Uzamış Yoğun Bakım Ünitesi Yatışı Arasındaki İlişki

İB Büşra GÜVENDİ ŞENGÖR¹, İB Mehmet ŞENGÖR², İB Muhammet Mucahit TİRYAKİ³, İB Faruk AKALIN¹, İB Kadir BIYIKLI¹, İB Mustafa Ferhat KETEN¹, İB Cemalettin YILMAZ³, İB Ünal AYDIN², İB Regayip ZEİR¹

¹University of Health Sciences Türkiye, Kartal Koşuyolu Education and Research Hospital, Department of Cardiology, İstanbul, Türkiye

²Bezmialem Vakıf University Faculty of Medicine, Department of Cardiovascular Surgery, İstanbul, Türkiye

³Yalova University Faculty of Medicine, Department of Cardiology, Yalova, Türkiye

ABSTRACT

Objective: Patients requiring longer postoperative intensive care unit (ICU) stay after coronary artery bypass grafting (CABG) are more likely to experience unfavorable clinical outcomes and increased healthcare resource utilization. The AHEAD score has been reported to predict adverse outcomes in several cardiovascular conditions. The present study sought to evaluate whether the AHEAD score could predict prolonged ICU stay in patients undergoing isolated CABG.

Methods: We analyzed data from 518 consecutive patients undergoing isolated CABG at two tertiary referral centers between January 2020 and September 2025. A postoperative ICU stay longer than 72 hours was considered prolonged. The AHEAD score, consisting of atrial fibrillation, low hemoglobin level, elderly age, abnormal renal function, and diabetes mellitus, was evaluated both as a continuous variable and according to predefined score categories (0, 1, 2, and ≥ 3 points).

Results: Of the 518 patients, 143 (27.6%) experienced prolonged ICU stay. These patients had significantly higher AHEAD scores than those without prolonged ICU stay [2 (1-3) vs. 1 (0-2), $p < 0.001$]. In multivariable analysis, increasing AHEAD score values were associated with a progressively

ÖZ

Amaç: Koroner arter bypass greftleme (KABG) sonrası uzamış yoğun bakım ünitesi (YBÜ) yatışı; olumsuz klinik sonuçlar ve artmış sağlık hizmeti maliyetleri ile ilişkilidir. AHEAD skorunun çeşitli kardiyovasküler hastalıklarda prognostik değeri gösterilmiştir. Bu çalışmada AHEAD skoru ile uzamış YBÜ yatışı arasındaki ilişkinin değerlendirilmesi amaçlandı.

Yöntemler: Bu retrospektif iki merkezli çalışmaya Ocak 2020-Eylül 2025 tarihleri arasında izole KABG uygulanan 518 hasta dahil edildi. Uzamış YBÜ yatışı, 72 saatin üzerindeki yatış süresi olarak tanımlandı. Atrial fibrilasyon, düşük hemoglobin düzeyi, ileri yaş, anormal böbrek fonksiyonu ve diyabetes mellitus bileşenlerinden oluşan AHEAD skoru hem sürekli değişken olarak hem de önceden tanımlanmış skor kategorilerine (0, 1, 2 ve ≥ 3 puan) göre analiz edildi.

Bulgular: Toplam 518 hastanın 143'ünde (%27,6) uzamış YBÜ yatışı saptandı. Uzamış YBÜ yatışı olan hastalarda AHEAD skoru anlamlı olarak daha yüksekti [2 (1-3) vs. 1 (0-2), $p < 0,001$]. Çok değişkenli analizde AHEAD skorundaki her 1 puanlık artış, uzamış YBÜ yatışı ile bağımsız olarak ilişkili bulundu [olasılık oranı (OO): 1,620; %95 güven aralığı (GA): 1,317-1,993; $p < 0,001$]. Uzamış YBÜ yatışı sıklığı AHEAD skor kategorileri boyunca kademeli olarak arttı ve skor ≥ 3 olan

Address for Correspondence: Büşra Güvendi Şengör, MD, University of Health Sciences Türkiye, Kartal Koşuyolu Education and Research Hospital, Department of Cardiology, İstanbul, Türkiye

E-mail: guvendi2861@hotmail.com

ORCID IDs of the authors: B.G.Ş.: 0000-0001-7946-1229, M.Ş.: 0000-0003-3253-2936, M.M.T.: 0000-0003-4882-9072, F.A.: 0009-0002-5357-7076, K.B.: 0000-0002-0841-1301, M.F.K.: 0000-0003-2646-4981, C.Y.: 0000-0003-4140-9139, Ü.A.: 0000-0001-6635-108X, R.Z.: 0000-0003-2729-884X

Cite this article as: Güvendi Şengör B, Şengör M, Tiryaki MM, Akalin F, Biyıklı K, Ketten MF, et al. AHEAD score and prolonged intensive care unit stay following isolated coronary artery bypass grafting. Bezmialem Science. 2026;14(4):385-392

Received: 29.06.2026

Accepted: 24.07.2026

Published date: 26.10.2026



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

ABSTRACT

greater likelihood of prolonged ICU stay [odds ratio (OR): 1.620, 95% confidence interval (CI): 1.317-1.993, $p<0.001$]. The frequency of prolonged ICU stays increased progressively across AHEAD score categories, from 16.3% in patients with a score of 0 to 61.2% in those with a score ≥ 3 . Compared with patients with an AHEAD score of 0, those with a score of 2 and ≥ 3 had significantly higher odds of prolonged ICU stay (OR: 2.060, 95% CI: 1.120-3.788, $p=0.020$ and OR: 8.121, 95% CI: 4.089-16.129, $p<0.001$, respectively).

Conclusion: The AHEAD score was independently associated with prolonged ICU stay following isolated CABG. Although it is straightforward to calculate using routinely available clinical variables, further studies are needed to establish its clinical role.

Keywords: Coronary artery bypass grafting, AHEAD score, intensive care unit

Öz

hastalarda %61,2'ye ulaştı. AHEAD skoru 0 olan hastalarla karşılaştırıldığında, skoru 2 ve ≥ 3 olan hastalarda uzamış YBÜ yatışı olasılığı anlamlı derecede daha yüksekti (sırasıyla OO: 2,060; %95 GA: 1,120-3,788; $p=0,020$ ve OO: 8,121; %95 GA: 4,089-16,129; $p<0,001$).

Sonuç: AHEAD skoru, izole KABG sonrası uzamış YBÜ yatışı ile bağımsız olarak ilişkilidir. Rutin olarak elde edilebilen klinik değişkenlere dayanan ve kolaylıkla hesaplanabilen bu skorun klinik rolünün belirlenebilmesi için ileri araştırmalara ihtiyaç vardır.

Anahtar Kelimeler: Koroner arter bypass greftleme, AHEAD skoru, yoğun bakım ünitesi

Introduction

Coronary artery bypass grafting (CABG) constitutes an established therapeutic option for patients with advanced coronary artery disease and remains widely performed across contemporary cardiac surgical procedures (1). In appropriately selected patients, CABG has been associated with improved survival, symptom relief, and better quality of life (2,3). Nevertheless, adverse postoperative outcomes continue to occur and contribute substantially to morbidity and mortality following surgery (4).

Several factors, including comorbidities, postoperative complications and prolonged intensive care unit (ICU) stay, have been associated with adverse outcomes after CABG (5,6). Among these factors, extended ICU stay is of particular importance due to increased healthcare resource utilization and worse postoperative outcomes (7). Therefore, reliable identification of individuals at increased risk before surgery may facilitate perioperative planning and optimize resource allocation. Although numerous risk models have been proposed for this purpose, there remains a need for simple and practical scoring systems that can be easily applied in routine clinical practice (8,9).

The AHEAD score is a straightforward clinical risk score derived from five readily available variables: atrial fibrillation (AF), anemia, advanced age, abnormal renal function and diabetes mellitus. Originally developed for patients hospitalized with acute heart failure, the score has subsequently shown prognostic relevance across several cardiovascular populations, including patients with coronary syndromes (10,11). However, its usefulness in the setting of CABG has not yet been explored. Accordingly, the present study aimed to evaluate the relationship between the AHEAD score and prolonged ICU stay following isolated CABG.

Methods

This study retrospectively reviewed 518 consecutive patients undergoing isolated CABG at two referral centers between January 2020 and September 2025. Patients undergoing urgent or emergent CABG, concomitant cardiac procedures, or redo cardiac surgery, as well as those with active infection or sepsis, malignancy, or end-stage kidney disease requiring dialysis, were excluded. Baseline clinical features, comorbidities, preoperative laboratory findings, operative variables, and postoperative outcomes were recorded. Relevant information was retrieved from electronic medical records and institutional databases maintained at both participating centers. Ethical approval was obtained from the University of Health Sciences Türkiye, Kartal Koşuyolu Education and Research Hospital Ethics (committee date: June 16, 2026; decision no: 2026/11/1506) and all study procedures complied with the principles outlined in the Declaration of Helsinki.

For each patient, the AHEAD score was preoperatively determined using five predefined clinical variables: AF, hemoglobin level <13 g/dL in men and <12 g/dL in women, elderly age (>70 years), abnormal renal function [serum creatinine >130 $\mu\text{mol/L}$ (approximately 1.47 mg/dL)], and diabetes mellitus. One point was awarded for each component, yielding a cumulative ranging from 0 to 5 (10). Intraoperative data regarding the number of distal anastomoses, cardiopulmonary bypass (CPB) time, and aortic cross-clamp (CC) time were recorded. Postoperative ICU stay and total hospital stay were also documented.

The study endpoint was prolonged ICU stay. Consistent with previous reports, patients requiring more than 72 hours of postoperative intensive care were classified as having prolonged ICU stay (6,9,12). For comparative analyses, the study population was divided into patients with ICU stays of ≤ 3 days and those requiring >3 days of

intensive care. Patients were also grouped according to predefined AHEAD score categories (0, 1, 2, and ≥ 3 points) to examine whether the likelihood of prolonged ICU stay increased across score levels. Furthermore, the relationship between the AHEAD score and the duration of intensive care utilization was evaluated.

Statistical Analysis

Statistical analyses were performed using Jamovi version 2.5.7.0 (The Jamovi project, Sydney, Australia). The normality of continuous variables was assessed using normality tests and distributional evaluation. Since most continuous variables were non-normally distributed, continuous variables were expressed as median (interquartile range). Categorical variables were presented as numbers and percentages. Comparisons between patients with ICU stay ≤ 3 days and >3 days were performed using the Mann-Whitney U test for continuous variables and the chi-square test or Fisher's exact test for categorical variables, as appropriate.

The binomial logistic regression analysis was performed to evaluate the relationship between the AHEAD score and prolonged ICU stay. In the primary analysis, the AHEAD score was modeled as a continuous variable, with effect estimates reported per 1-point increase. Three regression models were constructed. Model 1 included only the AHEAD score. Model 2 was adjusted for clinically relevant non-AHEAD variables, including female sex, serum albumin level, left ventricular ejection fraction (LVEF), and log10-transformed troponin level. Model 3 additionally incorporated procedural variables, including CPB time and CC time. Individual components of the AHEAD score

were not included in multivariable models to avoid overadjustment. Troponin levels were logarithmically transformed because of their right-skewed distribution.

The AHEAD score was also analyzed according to score categories. Because sparse data and unstable estimates were observed in the highest individual score strata, patients with AHEAD scores ≥ 3 were combined into a single high-risk category. AHEAD score 0 was used as the reference category for categorical comparisons. Odds ratios (ORs) with corresponding 95% confidence intervals (CIs) were calculated. In addition, because ICU stay duration was not normally distributed, the relationship between the AHEAD score and actual ICU stay duration was assessed using Spearman rank correlation analysis. A two-sided p-value <0.05 was considered statistically significant.

Results

The study cohort consisted of 518 patients (Figure 1). Among them, 375 (72.4%) patients had an ICU stay of ≤ 3 days, whereas 143 (27.6%) experienced prolonged ICU stay (>3 days). Baseline clinical characteristics according to prolonged ICU stay are presented in Table 1. Individuals in the prolonged ICU stay group were older [65 (57-71), $p=0.002$] and more frequently female (28.7%, $p=0.003$). In addition, diabetes mellitus, chronic kidney disease, and AF were also more prevalent among patients with prolonged ICU stay. Notably, the AHEAD score was significantly higher among patients with prolonged ICU stay than among those with ICU stay ≤ 3 days [2 (1-3) vs. 1 (0-2), $p<0.001$].

Laboratory and echocardiographic findings are summarized in Table 2. The prolonged-stay cohort exhibited lower

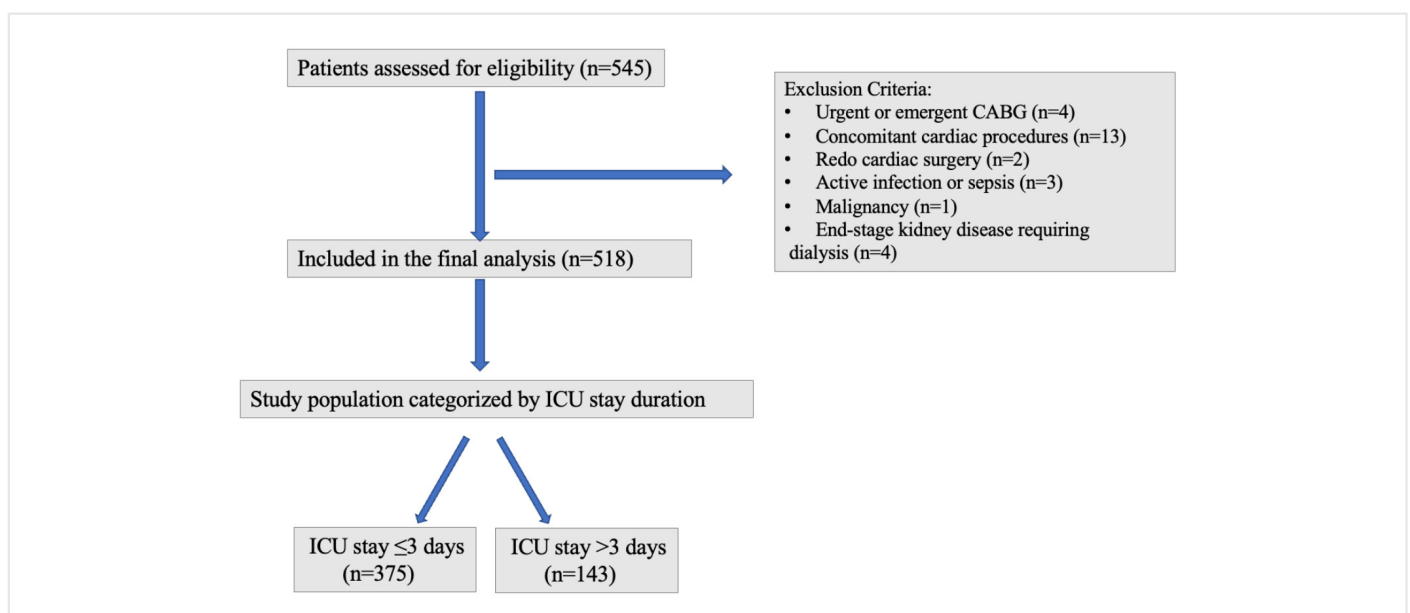


Figure 1. Study flow chart of patient selection and categorization

ICU: Intensive care unit, CABG: Coronary artery bypass grafting

hemoglobin and albumin levels [13.0 (11.5-14.4) vs. 13.6 (12.6-14.7) g/dL, $p<0.001$; 3.40 (3.10-3.95) vs. 3.80 (3.30-4.29), $p<0.001$; respectively]. In addition, neutrophil-to-lymphocyte ratio, creatinine, and troponin levels were

higher, whereas lymphocyte-to-monocyte ratio was lower, compared with patients with ICU stay ≤ 3 days [2.65 (1.83-3.57) vs. 2.21 (1.71-3.18), $p=0.022$; 0.96 (0.82-1.18) vs. 0.88 (0.80-1.05), $p=0.008$; 21.5 (3.30-1490) vs. 6.40 (0.72-122),

Table 1. Baseline clinical characteristics according to prolonged ICU stay

Variable	Overall	ICU stay ≤ 3 days n=375	ICU stay >3 days n=143	p-value
Age, years	62 [55-68]	61 [54-67]	65 [57-71]	0.002
Female sex	104 (20.1)	63 (16.8)	41 (28.7)	0.003
Height, cm	170 [165-175]	170 [165-175]	169 [160-173]	0.008
Weight, kg	80 [72-88]	80 [73-88]	78 [70.5-87]	0.219
BMI, kg/m ²	27.5 [25.3-30.2]	27.6 [25.4-30.1]	27.3 [25.2-30.9]	0.945
Obesity	115 (22.2)	75 (20)	40 (27.9)	0.289
Hypertension	373 (72.0)	266 (70.9)	107 (74.8)	0.378
Diabetes mellitus	271 (52.3)	178 (47.5)	93 (65.0)	<0.001
Recent myocardial infarction	170 (32.8)	121 (32.3)	49 (34.3)	0.665
Previous PCI	116 (22.4)	81 (21.6)	35 (24.5)	0.483
Chronic kidney disease	76 (14.7)	43 (11.5)	33 (23.1)	<0.001
Cerebrovascular disease	14 (2.7)	7 (1.9)	7 (4.9)	0.057
Hyperlipidemia	287 (55.4)	207 (55.2)	80 (55.9)	0.879
COPD	38 (7.3)	26 (6.9)	12 (8.4)	0.569
Atrial fibrillation	107 (20.7)	64 (17.1)	43 (30.1)	0.001
AHEAD score	1 [1-2]	1 [0-2]	2 [1-3]	<0.001

Values are presented as median [interquartile range] for continuous variables and n (%) for categorical variables
 BMI: Body mass index, PCI: Percutaneous coronary intervention, COPD: Chronic obstructive pulmonary disease, ICU: Intensive care unit

Table 2. Laboratory and echocardiographic findings according to prolonged ICU stay

Variable	Overall	ICU stay ≤ 3 days n=375	ICU stay >3 days n=143	p-value
WBC	8.36 [7.04-9.87]	8.28 [7.03-9.73]	8.58 [7.28-10.1]	0.153
Neutrophil count	5.13 [4.24-6.34]	5.10 [4.20-6.30]	5.43 [4.37-6.53]	0.148
Lymphocyte count	2.13 [1.60-2.68]	2.19 [1.70-2.66]	1.97 [1.46-2.75]	0.056
Monocyte count	0.63 [0.50-0.79]	0.63 [0.50-0.76]	0.64 [0.50-0.83]	0.584
Hemoglobin, g/dL	13.5 [12.2-14.6]	13.6 [12.6-14.7]	13.0 [11.5-14.4]	<0.001
Platelet count	227 [192-275]	227 [193-277]	220 [192-266]	0.325
NLR	2.30 [1.73-3.29]	2.21 [1.71-3.18]	2.65 [1.83-3.57]	0.022
LMR	3.58 [2.56-4.75]	3.64 [2.64-4.94]	3.25 [2.25-4.63]	0.045
Glucose	125 [100-173]	121 [99.3-167]	131 [101-192]	0.083
Creatinine	0.90 [0.81-1.07]	0.88 [0.80-1.05]	0.96 [0.82-1.18]	0.008
Urea	84.0 [66.0-94.0]	85.0 [70.0-95.0]	78.0 [57.0-91.0]	0.005
ALT	20.0 [14.0-29.0]	20.0 [14.0-29.0]	19.0 [13.0-28.5]	0.516
AST	20.0 [16.0-27.0]	20.0 [16.0-27.0]	20.0 [15.5-27.5]	0.928
Albumin	3.70 [3.30-4.20]	3.80 [3.30-4.29]	3.40 [3.10-3.95]	<0.001
Total cholesterol	194 [165-234]	199.5 [166-235]	189 [159-232.7]	0.198
LDL cholesterol	124 [95.8-156]	123 [95.9-156.2]	125 [94.9-155.7]	0.939
HbA1c	6.18 [5.61-7.68]	6.08 [5.59-7.40]	6.49 [5.73-7.78]	0.053
Troponin	8.60 [1.50-209]	6.40 [0.72-122]	21.5 [3.30-1490]	<0.001
LVEF, %	60 [50-60]	60 [50-60]	58 [49-60]	0.220

Values are presented as median [interquartile range]

WBC: White blood cell count, NLR: Neutrophil-to-lymphocyte ratio, LMR: Lymphocyte-to-monocyte ratio, ALT: Alanine aminotransferase, AST: Aspartate aminotransferase, HbA1c: Glycated hemoglobin, LVEF: Left ventricular ejection fraction, ICU: Intensive care unit, LDL: Low-density lipoprotein

$p < 0.001$, and 3.25 (2.25-4.63) vs. 3.64 (2.64-4.94), $p = 0.045$; respectively]. LVEF was comparable between the two groups. Procedural and hospitalization variables are shown in Table 3. CPB time was greater among individuals requiring prolonged ICU care [130 (110.8-148) vs. 118 (95-144) min, $p < 0.001$]. CC time and the number of distal anastomoses were comparable between the two groups. As expected, the prolonged ICU stay group had a significantly longer total hospital stay [14 (11-19.8) vs. 8 (7-12) days, $p < 0.001$].

The association between the AHEAD score and prolonged ICU stay is presented in Table 4. In the univariable model, each 1-point increase in AHEAD score was associated with higher odds of prolonged ICU stay (OR: 1.831, 95% CI: 1.506-2.227; $p < 0.001$). This observed association persisted after adjustment for clinically relevant non-AHEAD variables (Model 2: OR: 1.648, 95% CI: 1.343-2.021; $p < 0.001$), and after further adjustment for procedural variables (Model 3: OR: 1.620, 95% CI: 1.317-1.993; $p < 0.001$). Among the

covariates included in the multivariable models, lower albumin levels were also independently associated with prolonged ICU stay.

The distribution of prolonged ICU stays according to AHEAD score categories is presented in Table 5. The frequency of prolonged ICU stays increased progressively across AHEAD score categories: rising from 16.3% in patients with an AHEAD score of 0 to 61.2% in those with a score of ≥ 3 (Figure 2). Compared with patients with an AHEAD score of 0, those with an AHEAD score of 1 showed a numerically higher but statistically non-significant odds of prolonged ICU stay (OR: 1.501, 95% CI: 0.836-2.693; $p = 0.174$). Patients with an AHEAD score of 2 had approximately two-fold higher odds of prolonged ICU stay, whereas patients with an AHEAD score of ≥ 3 had more than eight-fold higher odds (OR: 2.060, 95% CI: 1.120-3.788; $p = 0.020$ and OR: 8.121, 95% CI: 4.089-16.129; $p < 0.001$, respectively). In a secondary analysis, the AHEAD score showed a weak yet statistically

Table 3. Procedural and hospitalization variables according to prolonged ICU stay

Variable	Overall n=518	ICU stay ≤ 3 days n=375	ICU stay > 3 days n=143	p-value
Cardiopulmonary bypass time, min	120 [100-145]	118 [95-144]	130 [110.8-148]	<0.001
Cross-clamp time, min	65 [52-80]	64 [52-80]	66.5 [56.8-78]	0.183
Distal anastomosis number	3 [3-4]	3 [3-4]	3 [3-4]	0.710
Beating-heart surgery	5 (1.0)	5 (1.3)	0 (0.0)	0.165
ICU stay, days	3 [2-4]	2 [2-3]	4 [4-6]	NA
Total hospital stays, days	10 [7-14]	8 [7-12]	14 [11-19.8]	<0.001

Values are presented as median [interquartile range] for continuous variables and n (%) for categorical variables
ICU: Intensive care unit, NA: Not applicable

Table 4. Logistic regression models for prolonged ICU stay

Variable	Model 1 OR (95% CI)	p-value	Model 2 OR (95% CI)	p-value	Model 3 OR (95% CI)	p-value
AHEAD score	1.831 (1.506-2.227)	<0.001	1.648 (1.343-2.021)	<0.001	1.620 (1.317-1.993)	<0.001
Female sex	—	—	1.481 (0.899-2.439)	0.123	1.496 (0.896-2.498)	0.123
Albumin	—	—	0.602 (0.409-0.885)	0.010	0.621 (0.420-0.917)	0.017
LVEF	—	—	0.995 (0.973-1.017)	0.654	0.996 (0.974-1.019)	0.725
Log10 troponin	—	—	1.119 (0.989-1.266)	0.074	1.117 (0.986-1.266)	0.083
Cardiopulmonary bypass time	—	—	—	—	1.002 (0.997-1.007)	0.501
Cross-clamp time	—	—	—	—	1.003 (0.991-1.016)	0.635
McFadden R ²	0.065		0.094		0.092	

Model 1 included only the AHEAD score. Model 2 was adjusted for clinically relevant non-AHEAD variables, including female sex, albumin, LVEF, and log10-transformed troponin. Model 3 additionally included procedural variables, including cardiopulmonary bypass time and cross-clamp time. Individual components of the AHEAD score were not included in the multivariable models to avoid overadjustment. Troponin was log10-transformed before inclusion in regression models because of its right-skewed distribution; therefore, the odds ratio for log10 troponin represents the change in odds per 10-fold increase in troponin level
OR: Odds ratio, CI: Confidence interval, LVEF: Left ventricular ejection fraction, ICU: Intensive care unit

significant positive correlation with postoperative ICU stay duration (Spearman’s rho: 0.283, $p<0.001$), indicating that higher AHEAD scores were associated with progressively longer ICU stays (Figure 3).

Discussion

In this multicenter cohort of patients undergoing isolated CABG, a higher AHEAD score was independently associated with an increased likelihood of prolonged ICU stay. The association remained significant after adjustment for relevant clinical, laboratory, and procedural variables. Moreover, the probability of prolonged ICU stay increased progressively across AHEAD score categories, indicating a clear risk gradient across increasing AHEAD score categories. Patients with an AHEAD score of 3 or greater represented a particularly vulnerable subgroup and experienced substantially higher rates of prolonged ICU stay than those with lower scores.

Prolonged ICU stay following CABG represents a clinically relevant outcome that reflects a more complicated postoperative course. Previous reports have demonstrated that patients requiring extended ICU stay are more likely to experience adverse events and consume greater healthcare

resources. Consequently, identifying patients at elevated risk before surgery may support individualized preoperative care and facilitate more efficient management of ICU care resources. Numerous prognostic models have been designed to facilitate outcome prediction after cardiac procedures (13-15). Dominici and colleagues constructed a nomogram including ten independent variables for the prediction of prolonged ICU stay (8). In the present study, increased age, female sex, diabetes mellitus, chronic kidney disease, and AF were more prevalent among patients with prolonged ICU stay.

The AHEAD score is a simple bedside tool based on five readily available clinical parameters: AF, anemia, advanced age, abnormal renal function, and diabetes mellitus. Although initially developed for patients hospitalized with acute heart failure, the AHEAD score has subsequently been investigated in several other cardiovascular settings, including patients with heart failure, coronary syndromes, and dementia (10,11,16,17). The present findings indicate that a higher AHEAD score was associated with prolonged ICU stay in patients undergoing CABG. The observed

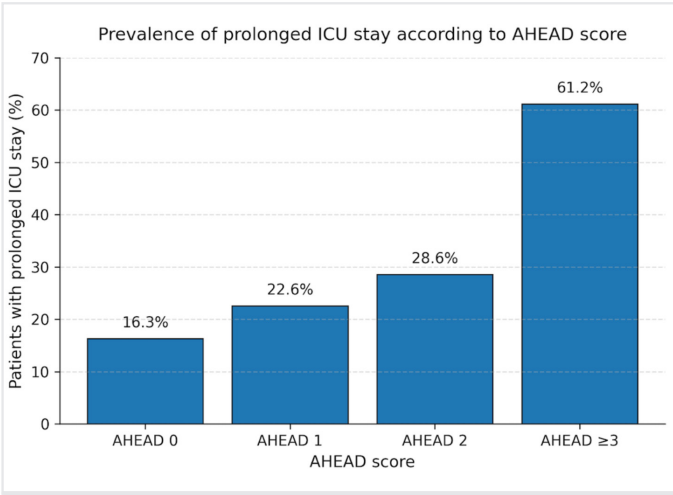


Figure 2. Prevalence of prolonged ICU stay according to AHEAD score categories
ICU: Intensive care unit

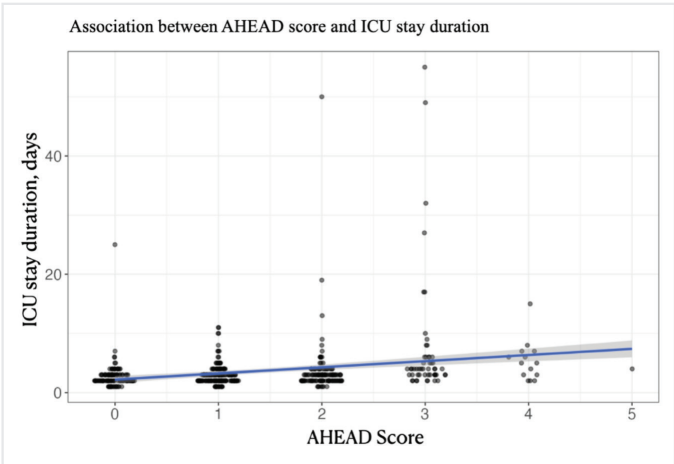


Figure 3. Association between AHEAD score and ICU stay duration. Scatter plot showing the relationship between AHEAD score and ICU stay duration. The solid blue line represents the fitted trend line with 95% confidence interval for visualization
ICU: Intensive care unit

Table 5. Prolonged ICU stay according to AHEAD score categories						
AHEAD score category	Overall n (%)	ICU stay ≤3 days n (%)	ICU stay >3 days n (%)	OR	95% CI	p-value
AHEAD 0	123 (23.7)	103 (83.7)	20 (16.3)	Reference	—	—
AHEAD 1	195 (37.6)	151 (77.4)	44 (22.6)	1.501	0.836-2.693	0.174
AHEAD 2	133 (25.7)	95 (71.4)	38 (28.6)	2.060	1.120-3.788	0.020
AHEAD ≥3	67 (12.9)	26 (38.8)	41 (61.2)	8.121	4.089-16.129	<0.001

Values are presented as n (%) unless otherwise indicated. Percentages for ICU stay ≤3 days and ICU stay >3 days are row percentages within each AHEAD score category. AHEAD 0 was used as the reference category for odds ratios. AHEAD scores ≥3 was combined into a single high-risk category because sparse data and unstable estimates were observed in the highest individual score strata
OR: Odds ratio, CI: Confidence interval, ICU: Intensive care unit

association between the AHEAD score and prolonged ICU stay is biologically plausible. Each component of the score has previously been linked to adverse postoperative outcomes following CABG. Age is an important parameter for postoperative recovery and advanced age is associated with reduced physiological reserve and increased susceptibility to postoperative complications (18,19). Diabetes mellitus is known to adversely affect postoperative recovery through impaired metabolic regulation and vulnerability to postoperative complications (20). Likewise, impaired renal function has consistently been linked to unfavorable outcomes after cardiac surgery, whereas preoperative anemia may compromise oxygen transport and increase perioperative transfusion requirements (21,22). In addition, AF often reflects a greater burden of cardiovascular disease and has been related to increased surgical risk and postoperative complications (23). The integration of these clinically relevant variables into a single score may account for the observed association between the AHEAD score and extended ICU utilization following CABG.

An additional finding of our study was the graded increase in prolonged ICU stay across AHEAD score categories. Patients with an AHEAD score ≥ 3 represented a particularly high-risk subgroup, with more than 60% experiencing prolonged ICU stay. This finding further supports the observed association between increasing AHEAD score categories and prolonged ICU stay. This graded relationship may be useful for future studies evaluating preoperative risk assessment in patients undergoing CABG.

Among the non-AHEAD variables included in the multivariable models, lower serum albumin levels were independently associated with prolonged ICU stay. Hypoalbuminemia is widely recognized as a marker of frailty, malnutrition, and systemic illness and has previously been linked to adverse outcomes after cardiac surgery (24,25). Our findings are consistent with these observations and further support the importance of preoperative nutritional and clinical status in determining postoperative recovery. While longer CC and CPB times have been linked to extended ICU utilization in previous reports, neither variable retained independent significance in our cohort (26,27). This discrepancy may be attributable to differences in patient characteristics between study populations as well as variations in perioperative management and ICU practices across centers. In contrast, the AHEAD score remained independently associated with prolonged ICU stay after multivariable adjustment.

Study Limitations

Several limitations merit consideration. First, the retrospective nature may have resulted in residual confounding despite multivariable adjustment. Second, because the study cohort was drawn from only two institutions, caution is warranted when extrapolating

these findings to other healthcare environments. Third, ICU discharge practices may vary between institutions and could influence ICU length of stay. Fourth, information regarding subsequent clinical events was unavailable, precluding assessment of the relationship between AHEAD score and outcomes beyond the immediate postoperative period. Finally, established surgical risk scores, including EuroSCORE II and ACEF, were not available in the retrospective dataset. Therefore, the present analysis could not adjust for these scores or compare the predictive performance of the AHEAD score with these established models. Further prospective studies are needed to determine whether the AHEAD score provides incremental prognostic information beyond established cardiac surgical risk scores.

Conclusion

A higher AHEAD score was independently associated with an increased likelihood of prolonged ICU stay after isolated CABG. Although the AHEAD score is simple to calculate using routinely available clinical variables, further prospective studies, including direct comparisons with established surgical risk models, are required to establish its clinical role.

Ethics

Ethics Committee Approval: Ethical approval was obtained from the University of Health Sciences Türkiye, Kartal Koşuyolu Education and Research Hospital Ethics (committee date: June 16, 2026; decision no: 2026/11/1506) and all study procedures complied with the principles outlined in the Declaration of Helsinki.

Informed Consent: The ethics committee waived the requirement for obtaining individual informed consent due to the retrospective nature of this research.

Declaration on the Use of Artificial Intelligence (AI) AI-based tools were employed exclusively for language refinement and grammatical editing purposes

Footnotes

Authorship Contributions

Surgical and Medical Practices: B.G.Ş., M.Ş., F.A., Ü.A., Concept: B.G.Ş., Ü.A., R.Z., Design: B.G.Ş., M.M.T., Ü.A., R.Z., Data Collection or Processing: M.Ş., M.M.T., F.A., K.B., M.F.K., C.Y., Analysis or Interpretation: B.G.Ş., M.Ş., M.M.T., F.A., K.B., M.F.K., C.Y., Ü.A., R.Z., Literature Search: B.G.Ş., M.Ş., F.A., K.B., M.F.K., C.Y., Writing: B.G.Ş., M.M.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.

Availability of Data and Material: The dataset analyzed during the current study can be obtained from the corresponding author upon reasonable request.

References

- Lawton JS, Tamis-Holland JE, Bangalore S, Bates ER, Beckie TM, Bischoff JM, et al. 2021 ACC/AHA/SCAI Guideline for coronary artery revascularization: executive summary: a report of the American College of Cardiology/American Heart Association Joint Committee on clinical practice guidelines. *Circulation*. 2022;145:e4-17.
- Creber RM, Dimagli A, Spadaccio C, Myers A, Moscarelli M, Demetres M, et al. Effect of coronary artery bypass grafting on quality of life: a meta-analysis of randomized trials. *Eur Heart J Qual Care Clin Outcomes*. 2022;8:259-68.
- Velazquez EJ, Lee KL, Jones RH, Al-Khalidi HR, Hill JA, Panza JA, et al.; STICHES Investigators. Coronary-artery bypass surgery in patients with ischemic cardiomyopathy. *N Engl J Med*. 2016;374:1511-20.
- Shroyer AL, Coombs LP, Peterson ED, Eiken MC, DeLong ER, Chen A, et al.; Society of Thoracic Surgeons. The Society of Thoracic Surgeons: 30-day operative mortality and morbidity risk models. *Ann Thorac Surg*. 2003;75:1856-64; discussion 1864-5.
- Chua TKT, Gao F, Chia SY, Sin KYK, Naik MJ, Tan TE, et al. Long-term mortality after isolated coronary artery bypass grafting and risk factors for mortality. *J Cardiothorac Surg*. 2024;19:429.
- Ibrahim KS, Kheirallah KA, Al Manasra ARA, Megdadi MA. Factors affecting duration of stay in the intensive care unit after coronary artery bypass surgery and its impact on in-hospital mortality: a retrospective study. *J Cardiothorac Surg*. 2024;19:45.
- Kao KD, Lee SKC, Liu CY, Chou NK. Risk factors associated with longer stays in cardiovascular surgical intensive care unit after CABG. *J Formos Med Assoc*. 2022;121(1 Pt 2):304-13.
- Dominici C, Salsano A, Nenna A, Spadaccio C, Barbato R, Mariscalco G, et al. A nomogram for predicting long length of stay in the intensive care unit in patients undergoing CABG: results from the multicenter E-CABG registry. *J Cardiothorac Vasc Anesth*. 2020;34:2951-61.
- Najafi M, Sheikhvatan M, Sheikhfathollahi M. Discriminative power of EuroSCORE in predicting morbidity and prolonged hospital stay in an Iranian sample population. *J Tehran Heart Cent*. 2014;9:15-9.
- Spinar J, Jarkovsky J, Spinarova L, Mebazaa A, Gayat E, Vitovec J, et al. AHEAD score—long-term risk classification in acute heart failure. *Int J Cardiol*. 2016;202:21-6.
- Tamaki S, Higaki A, Kawakami H, Nishimura K, Inoue K, Ikeda S, et al. Usefulness of the AHEAD score for prediction of all-cause death in patients with acute and chronic coronary syndromes. *Int J Cardiol Cardiovasc Risk Prev*. 2025;26:200457.
- Rustenbach CJ, Reichert S, Berger R, Schano J, Nemeth A, Haerberle H, et al. Unraveling the predictors for delirium and ICU stay duration in patients with heart failure and reduced ejection fraction (HFrEF) undergoing coronary artery bypass grafting—a multicentric analysis. *Biomedicine*. 2024;12:749.
- Cacciatore F, Belluomo Anello C, Ferrara N, Mazzella F, Manzi M, De Angelis U, et al. Determinants of prolonged intensive care unit stay after cardiac surgery in the elderly. *Aging Clin Exp Res*. 2012;24:627-34.
- Ettema RG, Peelen LM, Schuurmans MJ, Nierich AP, Kalkman CJ, Moons KG. Prediction models for prolonged intensive care unit stay after cardiac surgery: systematic review and validation study. *Circulation*. 2010;122:682-9, 7 p following p 689.
- Zhang P, Zhang X, Wang J, Zhang Y. Development of a prediction model for the risk of prolonged ICU length of stay in patients undergoing coronary artery bypass surgery: a retrospective analysis based on the MIMIC-IV database. *Medicine (Baltimore)*. 2026;105:e48224.
- Söylemez N, Toprak B, Karaca Ö, Yılmaz S, Ballı M, Ekici M, et al. Prognostic value of HALP and AHEAD scores for predicting 1-month heart failure following myocardial infarction. *J Clin Med*. 2026;15:4363.
- Hu WS, Lin CL. Comparison of CHA₂DS₂-VASc and AHEAD scores for the prediction of incident dementia in patients hospitalized for heart failure: a nationwide cohort study. *Intern Emerg Med*. 2019;14:395-402.
- Jafarkhani A, Imani B, Saeedi S, Shams A. Predictive factors of length of stay in intensive care unit after coronary artery bypass graft surgery based on machine learning methods. *Arch Acad Emerg Med*. 2025;13:e35.
- Nakasuji M, Matsushita M, Asada A. Risk factors for prolonged ICU stay in patients following coronary artery bypass grafting with a long duration of cardiopulmonary bypass. *J Anesth*. 2005;19:118-23.
- Huenges K, Rainer-Schmidt N, Panholzer B, Caliebe A, Hüttmann L, Kolat P, et al. Impact of diabetes in patients with acute myocardial infarction undergoing coronary artery bypass surgery within 48 hours. *Heart Lung Circ*. 2024;33:1272-9.
- Marui A, Okabayashi H, Komiya T, Tanaka S, Furukawa Y, Kita T, et al.; CREDO-Kyoto Investigators. Impact of occult renal impairment on early and late outcomes following coronary artery bypass grafting. *Interact Cardiovasc Thorac Surg*. 2013;17:638-43.
- Oprea AD, Del Rio JM, Cooter M, Green CL, Karhausen JA, Nailer P, et al. Pre- and postoperative anemia, acute kidney injury, and mortality after coronary artery bypass grafting surgery: a retrospective observational study. *Can J Anaesth*. 2018;65:46-59. English.
- Banach M, Mariscalco G, Ugurlucan M, Mikhailidis DP, Barylski M, Rysz J. The significance of preoperative atrial fibrillation in patients undergoing cardiac surgery: preoperative atrial fibrillation—still underestimated opponent. *Europace*. 2008;10:1266-70.
- Xu R, Hao M, Zhou W, Liu M, Wei Y, Xu J, et al. Preoperative hypoalbuminemia in patients undergoing cardiac surgery: a meta-analysis. *Surg Today*. 2023;53:861-72.
- Uzun N, Sever MY, Sarıdaş A, Yalınol M, Gülmez R, Atmaca S, et al. The Glasgow prognostic score (GPS) may serve as a prognostic indicator in ST-elevation myocardial infarction (STEMI) patients undergoing emergency coronary artery bypass graft surgery. *Bezmiâlem Science*. 2025;13:362-8.
- Madhavan S, Chan SP, Tan WC, Eng J, Li B, Luo HD, et al. Cardiopulmonary bypass time: every minute counts. *J Cardiovasc Surg (Torino)*. 2018;59:274-81.
- Ruggieri VG, Bounader K, Verhoye JP, Onorati F, Rubino AS, Gatti G, et al. Prognostic impact of prolonged cross-clamp time in coronary artery bypass grafting. *Heart Lung Circ*. 2018;27:1476-82.



Estimation of the Median Effective Dose of Propofol for Colonoscope Advancement to the Rectosigmoid Junction in Geriatric Patients: A Prospective Dose-finding Study

Geriatrik Hastalarda Kolonoskopun Rektosigmoid Bileşkeye İlerletilmesini Sağlayan Propofolün Medyan Etkili Dozunun Tahmini: Prospektif Doz Belirleme Çalışması

¹Bedirhan GÜNEL^{1,2}, ²Ayşe ŞENCAN¹

¹University of Health Sciences Türkiye, Kocaeli City Hospital, Department of Anesthesiology, Kocaeli, Türkiye

²Sakarya Training and Research Hospital, Clinic of Anesthesiology and Reanimation, Sakarya, Türkiye

ABSTRACT

Objective: This study aimed to determine the median effective dose (ED₅₀) of propofol monosedation required to achieve well-tolerated and motionless advancement of the colonoscope to the rectosigmoid junction in geriatric patients.

Methods: This prospective dose-finding study was conducted between September 8 and October 8, 2025. Geriatric patients aged ≥65 years with American Society of Anesthesiologists physical status I-III and body mass index <30 kg/m² who underwent elective colonoscopy under propofol monosedation were included. Propofol was administered using a modified Dixon up-and-down method, starting at 1.5 mg/kg with 0.1 mg/kg dose adjustments. Success was defined as colonoscope advancement to the rectosigmoid junction without movement, vocalization, or observable pain-related behavior, together with a Modified Observer's Assessment of Alertness/Sedation score of ≤2 at two minutes after induction. After each success, the next dose was decreased by 0.1 mg/kg; after each failure, it was increased by 0.1 mg/kg.

Results: Eight crossover points were observed, and the study concluded after 29 patients. The ED₅₀ of propofol was 0.57 mg/kg [95% confidence interval (CI), 0.357-0.791], and the ED₉₅ was 1.81 mg/kg (95% CI, 0.094-3.517). Mean total propofol dose was 128.4±47 mg. No clinically significant respiratory

Öz

Amaç: Bu çalışmada, geriatrik hastalarda işlemin iyi tolere edilmesini ve kolonoskopun hasta hareketi olmaksızın rektosigmoid bileşkeye ilerletilmesini sağlayan propofol monosedasyonunun medyan etkili dozunun (ED₅₀) belirlenmesi amaçlandı.

Yöntemler: Bu prospektif doz belirleme çalışması 8 Eylül-8 Ekim 2025 tarihleri arasında yürütüldü. Propofol monosedasyonu altında elektif kolonoskopi uygulanan, Amerikan Anesteziyologlar Derneği fiziksel durum sınıfı I-III olan, vücut kitle indeksi <30 kg/m² ve yaşı ≥65 yıl olan geriatrik hastalar çalışmaya dahil edildi. Propofol, 1,5 mg/kg başlangıç dozu ve 0,1 mg/kg doz aralıkları kullanılarak modifiye Dixon yukarı-aşağı yöntemiyle uygulandı. Başarı, indüksiyondan iki dakika sonra Modifiye Gözlemcinin Uyanıklık/Sedasyon Değerlendirme Ölçeği skorunun ≤2 olmasıyla birlikte kolonoskopun rektosigmoid bileşkeye hareket, vokalizasyon veya gözlenebilir ağrıyla ilişkili davranış olmaksızın ilerletilmesi olarak tanımlandı. Her başarılı yanıtın sonra bir sonraki hastada doz 0,1 mg/kg azaltıldı; her başarısız yanıtın sonra ise 0,1 mg/kg artırıldı.

Bulgular: Sekiz çapraz geçiş noktası gözlemlendi ve çalışma 29 hasta sonrasında tamamlandı. Propofolün medyan etkili dozu 0,57 mg/kg [%95 güven aralığı (GA), 0,357-0,791], %95 etkili

Address for Correspondence: Bedirhan Günel, MD, University of Health Sciences Türkiye, Kocaeli City Hospital, Department of Anesthesiology, Kocaeli; Sakarya Training and Research Hospital, Sakarya, Türkiye

E-mail: bedirhangunel71@gmail.com

ORCID IDs of the authors: B.G.: 0000-0003-2292-9403, A.Ş.: 0000-0003-2225-5269

Cite this article as: Günel B, Şencan A. Estimation of the median effective dose of propofol for colonoscope advancement to the rectosigmoid junction in geriatric patients: a prospective dose-finding study. Bezmialem Science. 2026;14(4):393-400

Received: 14.07.2026

Accepted: 12.08.2026

Published date: 26.10.2026



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

ABSTRACT

depression or airway intervention occurred; median lowest peripheral oxygen saturation (SpO₂) was 97% (91-99). Injection pain occurred in 20.7%, and post-procedural nausea in 3.4% of patients. Heart rate and SpO₂ did not change significantly. Although an overall difference in mean arterial pressure was observed (p=0.030), no significant pairwise differences remained after Bonferroni correction.

Conclusion: The ED₅₀ of propofol for well-tolerated and motionless colonoscopy insertion in geriatric patients was estimated to be 0.57 mg/kg. Given the exploratory nature of this dose-finding study, these findings should be interpreted as preliminary and may contribute to individualized dosing strategies for colonoscopy sedation in this population.

Keywords: Propofol, geriatric, geriatric anesthesia, effective dose, colonoscopy

Öz

dozu ise 1,81 mg/kg (%95 GA, 0,094-3,517) olarak hesaplandı. Ortalama toplam propofol dozu 128,4±47 mg idi. Klinik olarak anlamlı solunum depresyonu veya hava yolu müdahalesi gözlenmedi; en düşük periferik oksijen saturasyonu (SpO₂) medyan değeri %97 (91-99) idi. Enjeksiyon ağrısı hastaların %20,7'sinde, postoperatif bulantı ise %3,4'ünde görüldü. Kalp hızı ve SpO₂ değerlerinde anlamlı değişiklik gözlenmedi. Ortalama arter basıncında genel analizde anlamlı bir farklılık saptanmasına rağmen (p=0,030), Bonferroni düzeltmesi sonrası ikili karşılaştırmaların hiçbirinde anlamlı fark bulunmadı.

Sonuç: Geriatrik hastalarda iyi tolere edilen ve hareketsiz kolonoskop yerleştirilmesi için propofolün ED₅₀ değeri 0,57 mg/kg olarak tahmin edildi. Bu doz belirleme çalışmasının keşifsel niteliği göz önünde bulundurulduğunda, bulgular öncül sonuçlar olarak yorumlanmalıdır. Çalışmanın bulguları bu hasta popülasyonunda kolonoskopi sedasyonu için bireyselleştirilmiş dozlama stratejilerine katkı sağlayabilir.

Anahtar Kelimeler: Propofol, geriatri, geriatrik anestezi, etkili doz, kolonoskopi

Introduction

Propofol is one of the most commonly used agents for sedation during endoscopic procedures (1). It provides rapid emergence, a predictable and consistent depth of sedation, improved patient tolerance, and a complication profile comparable to or better than that of alternative agents (2-5). Propofol may be administered alone or in combination with other sedative agents, with the depth of sedation increasing in a dose-dependent manner (6). Its principal adverse effects include respiratory depression, dose-dependent hypotension, and injection pain, and no specific pharmacological antagonist is available (7,8). Although combination regimens may reduce the propofol dose required and thereby limit dose-dependent adverse effects, they may also delay recovery (6-8).

Age-related pharmacokinetic and pharmacodynamic changes affect the distribution of propofol and the electroencephalographic response to the drug (9). In geriatric patients, lower doses may be sufficient to achieve an equivalent depth of sedation because of age-related reductions in cardiorespiratory, hepatic, and renal function (10).

Nevertheless, evidence regarding the effective dose of propofol monosedation required for successful colonoscopy insertion in geriatric patients remains limited.

Therefore, the present study aimed to determine the median effective dose (ED₅₀) of propofol monosedation required for successful advancement of the colonoscopy to the rectosigmoid junction in adults aged ≥65 years.

Methods**Study Design**

A prospective dose-finding study was conducted using the up-and-down sequential allocation method. Ethical approval for the protocol was granted by the Kocaeli University Clinical Research Ethics Committee (approval no: KAEK/08.bl.02) on April 30, 2025, and the study conformed throughout to Declaration of Helsinki principles. Before the recruitment of participants, the trial was registered on ClinicalTrials.gov on August 25, 2025 (NCT07162805). Participant enrollment and data collection were carried out between September 8 and October 8, 2025. Written informed consent was obtained from every participant only after they had been fully briefed on the study procedures.

Participants

Adults aged ≥65 years, irrespective of sex, who had an American Society of Anesthesiologists (ASA) physical status of I-III and were scheduled to undergo colonoscopy were eligible for participation after providing written informed consent.

Patients were excluded if they declined to take part, had documented hypersensitivity to propofol, soy, peanuts, or egg white, were taking hypnotic, antidepressant, or antipsychotic drugs, or reported habitual alcohol use; had acute or chronic pain; and had either a body weight below 40 kg or a body mass index (BMI) above 30 kg/m²; had previously diagnosed obstructive sleep apnea or a STOP-Bang score of 3 or higher; had a history of anorectal disorders, including hemorrhoids, anal fissure, or anal fistula; had malignancy; had a history of colorectal surgery;

had chronic pulmonary disease or an active respiratory tract infection; had advanced renal or hepatic dysfunction; and had poorly controlled hypertension, defined as systolic blood pressure ≥ 180 mmHg and/or diastolic blood pressure ≥ 110 mmHg.

Randomization and Blinding

Due to the nature of the study, neither randomization nor blinding was performed.

Anesthetic Technique and Dose-finding Methodology

Standard monitoring, including pulse oximetry, electrocardiography, non-invasive blood pressure measurement, and capnography, was applied to every patient upon arrival in the endoscopy unit. Peripheral venous access was secured with a 20- or 22-gauge intravenous cannula, followed by administration of Ringer's lactate solution. Before sedation, patients received oxygen through a nasal cannula at 5 L/min for 3 minutes, and oxygen delivery at the same rate was continued until completion of the procedure. In line with standard clinical practice, colonoscopy was performed with the patient in the left lateral decubitus position. After completion of preoxygenation, intravenous propofol (2,6-diisopropylphenol; propofol 1%, Polifarma® Pharmaceutical Industry, Türkiye) was administered over 20 seconds to induce sedation.

The Dixon up-and-down sequential allocation method was used as the dose-finding strategy in this study (11,12). In this stepwise design, each participant's dose depends directly on how the previous participant responded—an approach long established in clinical pharmacology research (11,12). The starting dose was 1.5 mg/kg, selected with reference to previous endoscopy studies and routine clinical practice. Thereafter, the dose was increased or decreased by 0.1 mg/kg according to the outcome observed in the immediately preceding patient (13,14). Anesthetic management was performed by specialist anesthesiologists (B.G. and A.Ş.). After induction, a 2-minute interval was allowed to achieve maximum drug effect (15). At the beginning of the colonoscopy, the depth of sedation was assessed using the Modified Observer's Assessment of Alertness/Sedation (MOAA/S) scale (16). On this scale, a score of 5 indicates a prompt response to the patient's name spoken normally; 4, a sluggish response to the same stimulus; 3, a response only when the name is called loudly; 2, a response only to gentle-to-moderate physical prompting; and 1, a response elicited only by trapezius squeeze (16).

All procedures were conducted by a specialist endoscopist at the endoscopy unit of University of Health Sciences Türkiye, Kocaeli City Hospital, a tertiary-level referral center. Success was considered achieved when the colonoscope crossed the rectosigmoid junction without movement, vocalization, or observable pain-related behavior, together with a MOAA/S score of ≤ 2 at two minutes after induction.

If these criteria were not met—that is, if movement, vocalization, observable pain-related behavior, or a MOAA/S score of ≥ 3 occurred—the case was considered a failure, and a rescue dose of propofol (0.25-0.5 mg/kg) was titrated and administered. For the subsequent patient, the initial propofol dose was adjusted by 0.1 mg/kg, increased after failure and decreased after success.

Conditions for colonoscope insertion were evaluated only until the instrument had traversed the rectosigmoid junction. After the procedure began, rescue propofol was administered in titrated intravenous doses of 0.25-0.5 mg/kg whenever the patient exhibited observable pain-related behavior, moved, vocalized, or had a MOAA/S score of ≥ 3 . Sedation was subsequently adjusted to maintain a MOAA/S score of ≤ 2 throughout the procedure, with assessments performed at 2-minute intervals. Hemodynamic and respiratory variables, procedural duration, and total drug doses were documented. At the end of the examination, emergence was facilitated using verbal and tactile stimulation. Patients were then transferred to the post-anesthesia care unit and remained under observation until a modified Aldrete score of 10 was reached, after which discharge was permitted.

Primary Outcome

The primary outcome was the determination of the ED₅₀ of propofol required for successful colonoscope advancement to the rectosigmoid junction.

Secondary Outcomes

Secondary outcomes included calculation of the ED₉₅ of propofol required for successful colonoscope advancement to the rectosigmoid junction, the additional and total propofol doses administered, the presence of pain during injection, and the measurement of heart rate (HR), mean arterial pressure (MAP), and peripheral oxygen saturation (SpO₂) at four time points: before sedation (T0), during colonoscope passage through the rectosigmoid junction (T1), two minutes after insertion (T2), and at the end of the procedure (T3). Furthermore, the incidence of hypoxemia (SpO₂ $< 90\%$ for more than 10 seconds), the minimum SpO₂ value observed during the procedure, and procedure interruptions due to airway interventions (chin lift, jaw thrust, mask ventilation) were also evaluated as secondary outcome measures.

Sample Size

In a Dixon up-and-down sequential design study, the non-independence of data and the unknown distribution prevent the establishment of theoretically strict rules for sample size calculation (17). The sample size is determined according to the stop rule (17). Patients should continue to be enrolled until at least six crossover points between effective and ineffective responses are observed (17). Simulation studies have suggested that including 20-40 patients is sufficient to obtain stable estimates of the

target dose for most scenarios (17). Other anesthesia studies using this method have also typically included 20-40 patients (12). In the present study, eight crossover points were considered sufficient for this purpose, and the study was to be terminated once eight crossover points were reached.

Statistical Analysis

Statistical processing was carried out in IBM SPSS Statistics (version 27.0, IBM Corp., Armonk, NY). Normality of continuous variables was assessed using the Shapiro-Wilk test. Normally distributed data were presented as mean ± standard deviation, whereas non-normally distributed data were presented as median with range. Categorical variables were expressed as number and percentage. Temporal changes in hemodynamic and respiratory variables were analyzed using repeated-measures analysis of variance when parametric assumptions were met; otherwise, the Friedman test was used. Bonferroni adjustment was applied for subsequent pairwise comparisons. All tests were two-sided, and p-values <0.05 were considered statistically significant.

Dose-response analyses were performed separately in R software (R Foundation for Statistical Computing, Vienna, Austria). Binary success/failure responses were modeled against the natural logarithm of the initial propofol dose using logistic regression with a logit link (stats::glm, binomial

family). ED₅₀ and ED₉₅ values were obtained by inverse prediction from the fitted model coefficients. Standard errors were estimated using the delta method based on the variance-covariance matrix of the model coefficients, and 95% confidence intervals (CIs) were calculated on the original dose scale using the normal approximation. The fitted logistic regression model was also used to generate the dose-response curve.

Results

Participant Flow and Baseline Data

Recruitment was conducted between September 8 and October 8, 2025. Of the 51 patients assessed for eligibility, 22 did not meet the inclusion criteria, leaving 29 participants for enrollment (Figure 1). The study was terminated after the eighth crossover was reached. No enrolled participants were lost during the study, and all were included in the final analysis. The demographic and clinical characteristics of the study population are shown in Table 1, and the procedural data are presented in Table 2.

Primary Outcome

Analysis performed in R yielded a propofol ED₅₀ of 0.57 mg/kg (95% CI, 0.357-0.791 mg/kg) for successful colonoscope insertion. Responses were evaluated according to the modified Dixon up-and-down design, and the corresponding dose-response findings are presented in Figures 2 and 3.

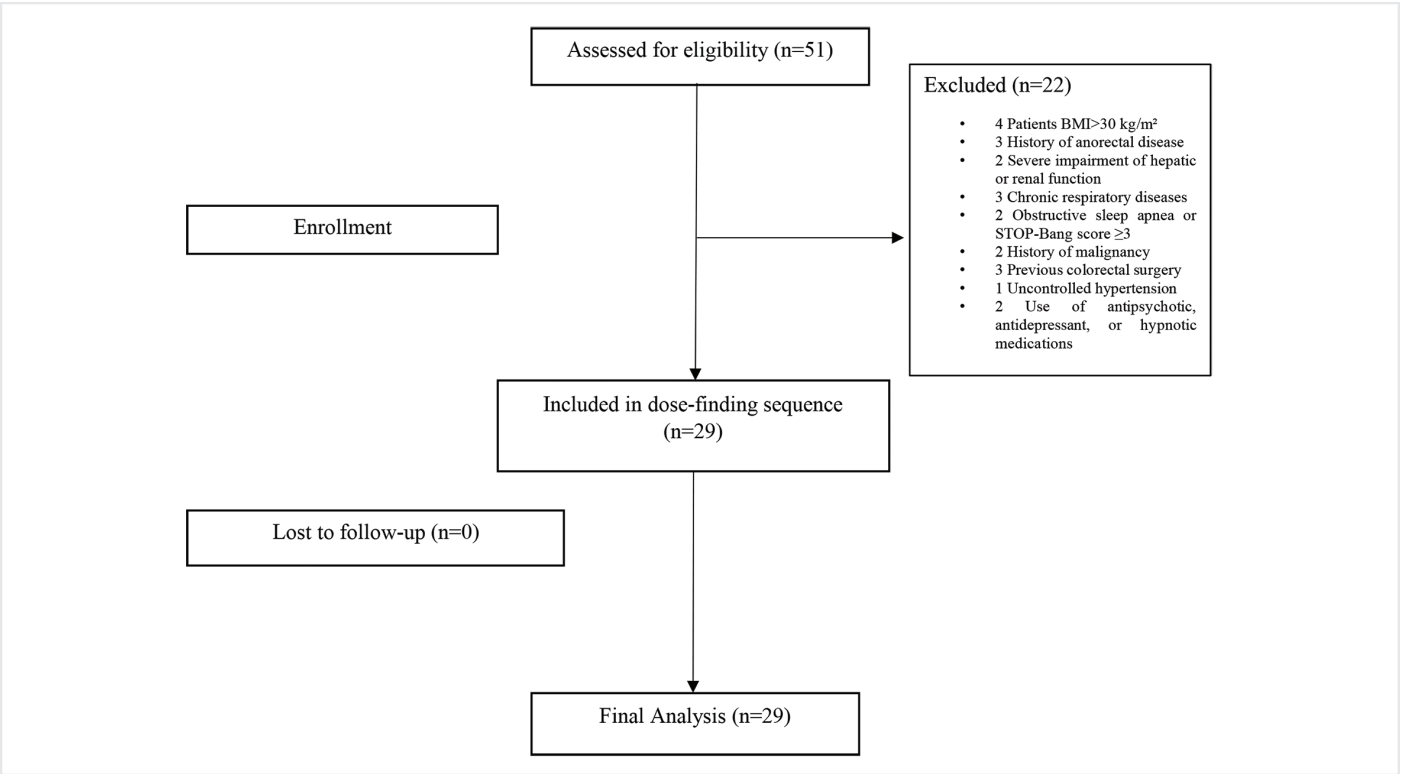


Figure 1. Flow-chart
BMI: Body mass index

Table 1. Patient baseline characteristics

Sex (male/female)	13 (44.8)/16 (55.2)
Age	69 (65-85)
Height (m)	1.7 (1.5-1.8)
Weight (kg)	74±10.1
BMI (kg/m ²)	25.2±3.1
Smoking	4 (13.8)
Hypertension	19 (65.5)
Diabetes mellitus	14 (48.3)
Cardiac disease	9 (31)
Other comorbidities	5 (17.2)
ASA classification II-III	13 (44.8)/16 (55.2)
Mallampati classification I-II	8 (27.6)/21 (72.4)
Indications for colonoscopy	
Surveillance	20 (69)
Suspected tumor	3 (10.3)
Diarrhea	2 (6.9)
Constipation	4 (13.8)
Data are presented as number (%), median (minimum-maximum) or mean ± standard deviation BMI: Body mass index, ASA: American Society of Anesthesiologists	

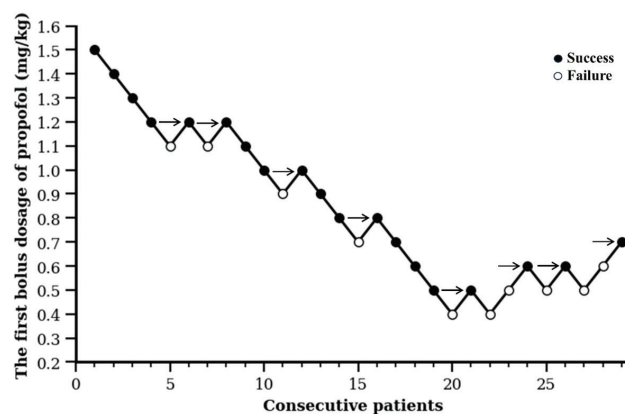
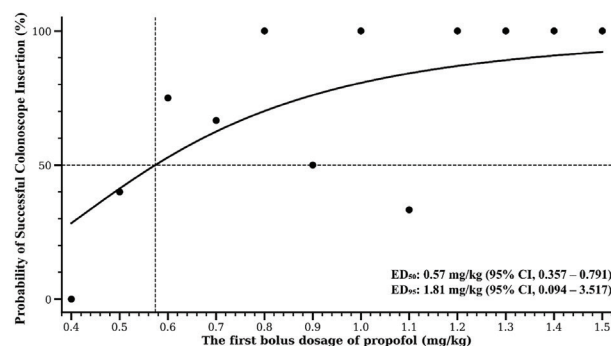
Table 2. Summary of perioperative variables

Initial bolus dose (mg)	61±21.4
Additional propofol dose (mg)	67.5±44.2
Total propofol dose (mg)	128.4±47
MOAA/S at 2 min post-induction (1/2/3/4)	3 (10.3)/21(72.4)/3 (10.3)/2 (6.9)
Procedure duration (min)	10 (4-24)
Lowest SpO ₂ (%)	97 (91-99)
Injection pain	6 (20.7)
Airway intervention	0 (0)
Post-procedural pain	1 (3.4)
Post-procedural nausea	1 (3.4)
Post-procedural vomiting	0 (0)
Data are presented as number (%), median (minimum-maximum) or mean ± standard deviation MOAA/S: Modified Observer's Assessment of Alertness/Sedation, SpO ₂ : Peripheral oxygen saturation	

Secondary Outcomes

Using R software, the ED₉₅ of propofol for successful colonoscope insertion was calculated as 1.81 mg/kg (95% CI, 0.094-3.517). Because this estimate exceeded the observed dose range and had a wide CI, it should be interpreted as an exploratory model-based estimate.

The total propofol dose administered during colonoscopy was 128.4±47 mg, with an additional post-induction dose of 67.5±44.2 mg (Table 2). The median procedure duration was 10 (4-24) min. Pain on injection after induction was

**Figure 2.** Responses of patients with the modified Dixon's up-and-down method**Figure 3.** Dose-response curve

The curve was derived from a logistic regression model relating binary success/failure responses to the natural logarithm of the initial propofol dose

CI: Confidence interval

observed in 6 patients (20.7%). The median of the lowest SpO₂ values recorded during the procedure was 97% (91-99%), and the incidence of airway intervention was 0 (0%). HR and SpO₂ measurements at the four time points did not differ significantly ($p>0.05$). An overall difference was observed in MAP across time points ($p=0.030$); however, none of the pairwise comparisons remained statistically significant after Bonferroni correction (Table 3).

Discussion

This study sought to determine the propofol dose associated with successful colonoscope insertion in geriatric patients with an ASA physical status of I-III and a BMI below 30 kg/m² who received propofol as the sole sedative agent. Based on the modified Dixon up-and-down design, the estimated ED₅₀ was 0.57 mg/kg (95% CI, 0.357-0.791 mg/kg). The

Table 3. Changes in hemodynamic and oxygenation parameters during the procedure

	Before sedation	Endoscope insertion	Two min after endoscope insertion	The end of the procedure	F	p-value
HR (beats/min)	73.3±8.3	72.3±7.5	75.6±9.3	74.7±9.7	2.011	0.160 ^A
MAP (mmHg)	93.2±12.7	86.7±13.4	87.9±14.5	89.1±13	3.648	0.030^{A*}
SpO ₂ (%)	98 (91-100)	98 (92-100)	97 (94-100)	98 (95-100)	-	0.129 ^F

Data are presented as median (minimum-maximum) or mean ± standard deviation

*: p<0.05 was considered statistically significant; since Bonferroni correction was applied in pairwise comparisons, no significant difference was found. ^A: Repeated-measures ANOVA, ^F: Friedman test

HR: Heart rate (beats per minute), MAP: Mean arterial pressure (mmHg), SpO₂: Peripheral oxygen saturation (%), ANOVA: Analysis of variance

corresponding ED₉₅ was estimated as 1.81 mg/kg (95% CI, 0.094-3.517 mg/kg). However, this estimate exceeded the range of doses actually observed in the study and had a wide CI. Therefore, the ED₉₅ should be regarded as a model-based exploratory extrapolation rather than a definitive clinical dosing threshold and requires validation in larger studies.

Rapid onset, short duration of effect, and a favorable recovery trajectory explain why propofol remains a preferred sedative for endoscopic work (1). Previous studies have demonstrated that propofol monotherapy does not present a significant disadvantage compared with combination therapy (4,18). Furthermore, reducing the propofol dose may help minimize clinically relevant dose-related adverse effects, including hypotension, respiratory depression, and excessive sedation (1). Because geriatric patients are more sensitive to propofol's effects on reducing systemic vascular resistance and causing myocardial depression, its use in this population requires special consideration (19). Although adverse effects are a concern in the geriatric population, current evidence suggests that propofol can be used safely at lower doses and with slow administration (4). Another important consideration during colonoscopy is that bowel preparation with sodium phosphate may reduce intravascular volume, thereby increasing the risk of hypotension when combined with sedation (3). Previous studies have investigated the effective dose of propofol in various patient populations (15,20-22). In a comparable study involving non-geriatric adults, Günel and Şencan (22) estimated the propofol ED₅₀ for successful colonoscopy insertion at 1.026 mg/kg (95% CI, 0.735-1.317 mg/kg). Liu et al. (15) reported propofol ED₅₀ and ED₉₅ values of 1.90 and 2.15 mg/kg, respectively, in healthy Chinese adults undergoing upper gastrointestinal endoscopy. In their study, the effective dose was estimated using the modified Dixon up-and-down method, in which each patient's dose was adjusted according to the response of the previous patient (15,21). This method typically requires a minimum of six crossover points to provide reliable results (15). Similar to the study by Liu et al. (15), patient enrollment in our trial was discontinued after eight crossover points were obtained. Another similarity with their study was the use of the MOAA/S score to assess the depth of sedation; in our study, we targeted a MOAA/S score of ≤2 (15). Notably, the ED₅₀ and ED₉₅ values in our

study were markedly lower than those reported by Liu et al. (15), which can largely be attributed to the lower propofol requirements needed to achieve the same depth of sedation in geriatric patients. Hayes et al. (20) also reported an ED₅₀ of 6.1 mg/kg for propofol during deep sedation for gastroduodenoscopy in children. Although this value is considerably higher than that observed in our study, it is not surprising, as children have a larger volume of distribution for propofol, necessitating higher doses to achieve effective sedation (23).

A recent meta-analysis by Ahmer et al. (24) evaluated randomized controlled trials comparing propofol and remimazolam in geriatric patients undergoing gastrointestinal endoscopic procedures and found that propofol was superior in terms of sedation success after the initial dose and time to loss of consciousness. Remimazolam reduced the rates of bradycardia, hypoxemia, and injection pain; however, procedure time, recovery and discharge intervals, respiratory depression, and hypotension were comparable between groups. The authors therefore considered propofol a suitable and effective sedative for geriatric patients (24). Therefore, in order to ensure successful colonoscopy insertion while avoiding excessive sedation, we evaluated the ED₅₀ and ED₉₅ of propofol, still the most commonly used agent for gastrointestinal procedures, in monosedation in the geriatric population. We believe that our findings will contribute valuable data to the literature and inform clinical practice.

By contrast, Borrat et al. (25) reported that 0.75-1 mg/kg of propofol was sufficient for successful upper gastrointestinal endoscopy in patients older than 60 years. Although their patients had demographic characteristics similar to those in our study, the effective dose was slightly higher, possibly because they targeted a MOAA/S score of ≥3 and administered propofol with remifentanyl using target-controlled infusion, whereas we used propofol boluses alone (25). Ethnic differences may also have contributed to the discrepancy in effective dose (15,25).

Injection pain occurred in six patients (20.7%). This rate was lower than that generally reported in the literature (26), but higher than the 11.5% incidence reported by Liu et al. (15). Differences in venous access site, injection rate, and sample size may partly explain the variability among studies.

Study Limitations

Several limitations warrant mention. All enrolled patients were 65 years or older with a BMI under 30 kg/m², and although ASA I-III patients were eligible, the final cohort consisted only of ASA II-III individuals. Therefore, the drug regimen used and the results obtained may not be generalizable to other age groups or different patient populations. Second, sedation depth was assessed solely using the MOAA/S score rather than objective measures such as the bispectral index or electroencephalography. Third, plasma and/or effect-site propofol concentrations were not measured; therefore, a direct dose-exposure-response relationship could not be established, and the findings were based solely on clinical outcomes. Fourth, the limited sample size and exploratory nature of this dose-finding study should be considered when interpreting the dose estimates. In particular, although the modified Dixon up-and-down method is suitable for estimating the ED₅₀, observations tend to cluster around the ED₅₀, limiting the reliability of estimates at the upper end of the dose-response curve, such as the ED₉₀ and ED₉₅. Accordingly, the wide CI of the ED₉₅ estimate in the present study reflects this methodological limitation, and the ED₉₅ should be regarded as exploratory and hypothesis-generating rather than as a definitive clinical dosing recommendation. Fifth, the secondary outcomes evaluated in this cohort should be interpreted cautiously because of the limited sample size. In particular, the absence of clinically significant adverse events or airway interventions should not be regarded as definitive evidence of safety, but rather as an exploratory finding that requires confirmation in larger studies.

Conclusion

Application of the modified Dixon up-and-down approach yielded an estimated propofol ED₅₀ of 0.57 mg/kg (95% CI, 0.357-0.791) for successful passage of the colonoscope through the rectosigmoid junction in geriatric patients. This estimate may serve as an initial reference for tailoring propofol administration during colonoscopy in geriatric patients. However, confirmation in larger and more diverse cohorts is required before a more definitive dosing recommendation can be established.

Ethics

Ethics Committee Approval: Ethical approval for the protocol was granted by the Kocaeli University Clinical Research Ethics Committee (approval no: KAEK/08.bl.02) on April 30, 2025, and the study conformed throughout to Declaration of Helsinki principles.

Informed Consent: Written informed consent was obtained from every participant only after they had been fully briefed on the study procedures.

Declaration on the Use of Artificial Intelligence (AI)

For transparency, the authors note that an artificial intelligence-assisted language model was utilized for 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

Surgical and Medical Practices: B.G., A.Ş., Concept: B.G., Design: B.G., Data Collection or Processing: B.G., A.Ş., Analysis or Interpretation: B.G., Literature Search: B.G., A.Ş., Writing: B.G., A.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.

References

1. Yoon SW, Choi GJ, Lee OH, Yoon IJ, Kang H, Baek CW, et al. Comparison of propofol monotherapy and propofol combination therapy for sedation during gastrointestinal endoscopy: a systematic review and meta-analysis. *Dig Endosc.* 2018;30:580-91.
2. Delgado AAA, de Moura DTH, Ribeiro IB, Bazarbashi AN, Dos Santos MEL, Bernardo WM, et al. Propofol vs traditional sedatives for sedation in endoscopy: a systematic review and meta-analysis. *World J Gastrointest Endosc.* 2019;11:573-88.
3. García Guzzo ME, Fernandez MS, Sanchez Novas D, Salgado SS, Terrasa SA, Domenech G, et al. Deep sedation using propofol target-controlled infusion for gastrointestinal endoscopic procedures: a retrospective cohort study. *BMC Anesthesiol.* 2020;20:195.
4. Kerker A, Hardt C, Schlieff HE, Dumoulin FL. Combined sedation with midazolam/propofol for gastrointestinal endoscopy in elderly patients. *BMC Gastroenterol.* 2010;10:11.
5. Wang D, Chen C, Chen J, Xu Y, Wang L, Zhu Z, et al. The use of propofol as a sedative agent in gastrointestinal endoscopy: a meta-analysis. *PLoS One.* 2013;8:e53311.
6. Riphaus A, Geist C, Schrader K, Martchenko K, Wehrmann T. Intermittent manually controlled versus continuous infusion of propofol for deep sedation during interventional endoscopy: a prospective randomized trial. *Scand J Gastroenterol.* 2012;47:1078-85.
7. Paspatis GA, Manolaraki MM, Vardas E, Theodoropoulou A, Chlouverakis G. Deep sedation for endoscopic retrograde cholangiopancreatography: intravenous propofol alone versus intravenous propofol with oral midazolam premedication. *Endoscopy.* 2008;40:308-13.
8. Hsieh YH, Chou AL, Lai YY, Chen BS, Sia SL, Chen IC, et al. Propofol alone versus propofol in combination with meperidine for sedation during colonoscopy. *J Clin Gastroenterol.* 2009;43:753-7.
9. Kim Y, Park J, Chung W, Jo Y, Oh C, Hong B, et al. Age-related electroencephalographic delta and alpha oscillations during

- sedation with target-controlled propofol infusion. *J Clin Med*. 2025;14:3024.
10. Purdon PL, Pavone KJ, Akeju O, Smith AC, Sampson AL, Lee J, et al. The ageing brain: age-dependent changes in the electroencephalogram during propofol and sevoflurane general anaesthesia. *Br J Anaesth*. 2015;115(Suppl 1):i46-57.
11. Dixon WJ. Staircase bioassay: the up-and-down method. *Neurosci Biobehav Rev*. 1991;15:47-50.
12. Pace NL, Stylianou MP. Advances in and limitations of up-and-down methodology: a précis of clinical use, study design, and dose estimation in anesthesia research. *Anesthesiology*. 2007;107:144-52.
13. Tang L, Lv W, Sun J, Chen L. Effect of intravenous lidocaine on the ED₅₀ of propofol induction dose in elderly patients undergoing painless gastroscopy: a prospective, randomized, double-blinded, controlled study. *Perioper Med (Lond)*. 2024;13:15.
14. Hu S, Wang M, Li S, Zhou W, Zhang Y, Shi H, et al. Intravenous lidocaine significantly reduces the propofol dose in elderly patients undergoing gastroscopy: a randomized controlled trial. *Drug Des Devel Ther*. 2022;16:2695-705.
15. Liu FK, Wan L, Shao LJZ, Zou Y, Liu SH, Xue FS. Estimation of effective dose of propofol mono-sedation for successful insertion of upper gastrointestinal endoscope in healthy, non-obese Chinese adults. *J Clin Pharm Ther*. 2021;46:484-91.
16. Cohen LB, Delegge MH, Aisenberg J, Brill JV, Inadomi JM, Kochman ML, et al.; AGA Institute. AGA Institute review of endoscopic sedation. *Gastroenterology*. 2007;133:675-701.
17. Zhang J, Kong L, Ni J. ED₅₀ and ED₉₅ of propofol combined with different doses of intravenous lidocaine for first-trimester uterine aspiration: a prospective dose-finding study using up-and-Down sequential allocation method. *Drug Des Devel Ther*. 2022;16:3343-52.
18. Singh H, Poluha W, Cheung M, Choptain N, Baron KI, Taback SP. Propofol for sedation during colonoscopy. *Cochrane Database Syst Rev*. 2008;2008:CD006268. Update in: *Cochrane Database Syst Rev*. 2025;10:CD006268.
19. Phillips AT, Deiner S, Mo Lin H, Andreopoulos E, Silverstein J, Levin MA. Propofol use in the elderly population: prevalence of overdose and association with 30-day mortality. *Clin Ther*. 2015;37:2676-85.
20. Hayes J, Matava C, Pehora C, El-Beheiry H, Jarvis S, Finkelstein Y. Determination of the median effective dose of propofol in combination with different doses of ketamine during gastro-duodenoscopy in children: a randomised controlled trial. *Br J Anaesth*. 2018;121:453-61.
21. Li S, Yu F, Zhu H, Yang Y, Yang L, Lian J. The median effective concentration (EC₅₀) of propofol with different doses of fentanyl during colonoscopy in elderly patients. *BMC Anesthesiol*. 2016;16:24.
22. Gunel B, Sencan A. Estimation of the median effective dose of propofol for colonoscope insertion in adult patients: a prospective dose-finding study. *J Cukurova Anesth Surg*. 2026;9:378-89.
23. Marsh B, White M, Morton N, Kenny GN. Pharmacokinetic model driven infusion of propofol in children. *Br J Anaesth*. 1991;67:41-8.
24. Ahmer W, Imtiaz S, Alam DM, Ahmed K, Sajid B, Yousuf J, et al. Remimazolam versus propofol for sedation in gastrointestinal endoscopy and colonoscopy within elderly patients: a meta-analysis of randomized controlled trials. *Eur J Clin Pharmacol*. 2024;80:493-503.
25. Borrat X, Valencia JF, Magrans R, Gimenez-Mila M, Mellado R, Sendino O, et al. Sedation-analgesia with propofol and remifentanyl: concentrations required to avoid gag reflex in upper gastrointestinal endoscopy. *Anesth Analg*. 2015;121:90-6.
26. Wang W, Zhou L, Wu LX, Wang T, Zhang CB, Sun L. 5-HT₃ receptor antagonists for propofol injection pain: a meta-analysis of randomized controlled trials. *Clin Drug Investig*. 2016;36:243-53.



Prognostic Value of Admission Red Cell Distribution Width and Mean Platelet Volume in Pediatric Sepsis and Septic Shock: Comparison with CRP, Procalcitonin, Lactate, PRISM IV and pSOFA

Pediatric Sepsis ve Septik Şokta Başvuru Sırasındaki Eritrosit Dağılım Genişliği ve Ortalama Trombosit Hacminin Prognostik Değeri: CRP, Prokalsitonin, Laktat, PRISM IV ve pSOFA ile Karşılaştırma

¹ Ferhat SARI¹, ² Gökhan CEYLAN², ³ Özlem SARAÇ³, ⁴ Gülhan ATAKUL³, ⁵ Ekin SOYDAN⁴, ⁶ Hasan AĞIN³

¹Bezmialem Vakıf University Faculty of Medicine, Department of Child Health and Diseases, Pediatric Intensive Care Unit, İstanbul, Türkiye

²Hamilton Medical AG Bonaduz, Department of Research and New Technology, Bonaduz, Switzerland

³University of Health Sciences Türkiye, Dr. Behçet Uz Pediatric Diseases and Surgery Training and Research Hospital, İzmir, Türkiye

⁴Aydın Maternity and Children's Hospital, Pediatric Intensive Care Unit, Aydın, Türkiye

ABSTRACT

Objective: In this study, we aimed to evaluate the relationship between red cell distribution width (RDW) and mean platelet volume (MPV) levels and disease severity in pediatric patients followed in the pediatric intensive care unit with a diagnosis of sepsis or septic shock, to investigate the correlations of these biomarkers with procalcitonin, C-reactive protein (CRP), lactate, PRISM, and pSOFA scores, and to assess their diagnostic performance.

Methods: This retrospective observational study included 243 pediatric patients aged 1 month to 18 years admitted to the pediatric intensive care unit with sepsis or septic shock. RDW, MPV, procalcitonin, CRP, lactate, PRISM, and pSOFA scores obtained at admission were analyzed. Non-parametric tests were used for group comparisons. Receiver operating characteristic analysis for 28-day mortality, and age-based subgroup analyses, were performed.

ÖZ

Amaç: Bu çalışmada sepsis ve septik şok tanısı ile çocuk yoğun bakımda izlenen çocuk hastalarda eritrosit dağılım genişliği (RDW) ve ortalama trombosit hacmi (MPV) düzeylerinin hastalık şiddeti ile ilişkisini değerlendirmek, bu biyobelirteçlerin prokalsitonin, C-reaktif protein (CRP), laktat, PRISM ve pSOFA skorları ile korelasyonunu incelemek ve tanılarda performanslarını değerlendirmeyi amaçladık.

Yöntemler: Bu retrospektif gözlemsel çalışmaya çocuk yoğun bakım ünitesinde sepsis veya septik şok tanısı ile izlenen, 1 ay-18 yaş arası toplam 243 hasta dahil edildi. Hastaların başvuru anındaki RDW, MPV, prokalsitonin, CRP ve laktat değerleri ile PRISM ve pSOFA skorları kaydedildi. Gruplar arası karşılaştırmalar non-parametrik testler ile yapıldı. Biyobelirteçlerin 28 günlük mortaliteyi öngörmedeki performansı alıcı işletim karakteristiği analizi ile değerlendirildi ve yaşa göre alt grup analizleri yapıldı.

Address for Correspondence: Ferhat Sarı, MD, Bezmialem Vakıf University Faculty of Medicine, Department of Child Health and Diseases, Pediatric Intensive Care Unit, İstanbul, Türkiye

E-mail: ferhatsari_dr@hotmail.com

ORCID IDs of the authors: F.S.: 0000-0002-9134-767X, G.C.: 0000-0002-1730-6968, Ö.S.: 0000-0003-2684-0625, G.A.: 0000-0002-3832-9691, E.S.: 0000-0003-2626-5499, H.A.: 0000-0003-3306-8899

Cite this article as: Sarı F, Ceylan G, Saraç Ö, Atakul G, Soydan E, Ağin H. Prognostic value of admission red cell distribution width and mean platelet volume in pediatric sepsis and septic shock: comparison with CRP, procalcitonin, lactate, PRISM IV and pSOFA. Bezmialem Science. 2026;14(4):401-409

Received: 02.07.2026

Accepted: 17.09.2026

Published date: 26.10.2026



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

ABSTRACT

Results: RDW, CRP, MPV, procalcitonin, and lactate levels were significantly higher in the septic shock group than the sepsis group ($p<0.001$). All biomarkers showed positive correlations with PRISM and pSOFA scores. For predicting 28-day mortality, lactate [area under the curve (AUC): 0.99] and CRP (AUC: 0.98) showed the highest discriminative power, followed by pSOFA (0.89), procalcitonin (0.87), and RDW (0.85), while MPV (0.73) showed the lowest performance. The combined RDW+CRP+MPV model improved discrimination (AUC: 0.99). CRP performed relatively better in infants (<12 months), while RDW and MPV performed relatively better in children aged ≥ 12 months.

Conclusion: RDW, CRP, MPV, procalcitonin, and lactate are readily available, cost-effective biomarkers associated with disease severity and mortality in pediatric sepsis and septic shock. CRP and lactate alone show strong performance, while RDW appears to be a more consistent prognostic marker than MPV. Their combined, age-specific use may support early risk stratification in clinical practice.

Keywords: Pediatric sepsis, RDW, PRISM IV, pSOFA

ÖZ

Bulgular: Septik şok grubunda RDW, CRP, MPV, prokalsitonin ve laktat düzeyleri sepsis grubuna kıyasla anlamlı olarak daha yüksek bulundu ($p<0,001$). Tüm biyobelirteçler PRISM ve pSOFA skorları ile pozitif korelasyon gösterdi (en güçlüsü prokalsitonin-pSOFA, $r=0,87$). Yirmi sekiz günlük mortaliteyi öngörmeye laktat [eğri altında kalan alan (AUC): 0,99] ve CRP (AUC: 0,98) en yüksek ayırt edici güce sahipti; bunu pSOFA (0,89), prokalsitonin (0,87) ve RDW (0,85) izledi; MPV (0,73) en düşük performansı gösterdi. RDW, CRP ve MPV'nin kombine modeli ayırt edici gücü artırdı (AUC: 0,99). <12 ay grubunda CRP'nin, ≥ 12 ay grubunda ise RDW ve MPV'nin görece daha güçlü performans gösterdiği saptandı.

Sonuç: RDW, CRP, MPV, prokalsitonin ve laktat pediatrik sepsis ve septik şokta hastalık şiddeti ve mortalite ile ilişkili, kolay erişilebilir ve düşük maliyetli biyobelirteçler olarak kullanılabilir. CRP ve laktatın tek başına güçlü prognostik belirteçler olduğu, RDW'nin ise MPV'ye kıyasla daha tutarlı bir prognostik belirteç olduğu görülmektedir; bu parametrelerin birlikte ve yaşa özgü kestirim değerleriyle kullanımı, erken risk sınıflamasında klinik olarak yararlı olabilir.

Anahtar Kelimeler: Pediatrik sepsis, RDW, PRISM IV, pSOFA

Introduction

Sepsis and septic shock remain major causes of morbidity and mortality in pediatric intensive care units (PICUs). Early recognition and accurate risk stratification constitute the cornerstone of clinical management. The clinical course of pediatric sepsis is highly variable and may rapidly progress to multiple organ dysfunction syndrome. Therefore, accurate assessment of disease severity at the time of intensive care unit admission is of critical importance.

Several scoring systems have been developed to assess disease severity and prognosis in pediatric critical care practice (1). The Pediatric Risk of Mortality IV (PRISM IV) score is a widely used prognostic tool for predicting mortality risk. The Pediatric Sequential Organ Failure Assessment (pSOFA) score was adapted to evaluate organ dysfunction and is considered a valid instrument for predicting disease severity and adverse outcomes in children with sepsis (2). However, the calculation of these scoring systems requires numerous clinical and laboratory variables and may be time-consuming in routine clinical practice.

In recent years, increasing attention has been directed toward the prognostic value of biomarkers that can be readily obtained from routine laboratory tests in patients with sepsis and septic shock. C-reactive protein (CRP), a classical acute-phase reactant, reflects the degree of infection and systemic inflammation. Previous studies have demonstrated that elevated CRP levels are associated with increased disease severity and unfavorable clinical outcomes in patients with sepsis and septic shock (3). Red cell distribution width (RDW), a parameter reflecting the heterogeneity of erythrocyte volume distribution, has been

reported to be associated with inflammation, oxidative stress, and bone marrow dysfunction (4). Although growing evidence suggests that RDW is associated with mortality and disease severity in adult intensive care and sepsis populations, data regarding its prognostic significance in pediatric patients remain limited. Mean platelet volume (MPV), which reflects platelet size and activity, is considered an indirect marker of platelet activation during inflammatory processes. Given the established role of platelets in inflammation, microthrombosis, and endothelial injury during the pathophysiology of sepsis, the clinical relevance of MPV has gained increasing attention (5).

Although the prognostic roles of RDW, CRP, and MPV in sepsis and septic shock have been investigated individually, studies evaluating these biomarkers in conjunction with PRISM and pSOFA scores and incorporating age-specific subgroup analyses in pediatric populations remain scarce.

The aim of this study was to evaluate the association of RDW, CRP, and MPV with disease severity in pediatric patients diagnosed with sepsis and septic shock, to investigate their correlations with PRISM and pSOFA scores, and to determine their performance in predicting the mortality using receiver operating characteristic (ROC) curve analysis.

Methods**Study Design and Patient Selection**

This retrospective observational study was conducted in a tertiary PICU using medical records from October 2020 to November 2025. During the study period, 1,440 patients

were admitted to the PICU, and 243 pediatric patients aged 1 month to 18 years who met the eligibility criteria were included. Demographic characteristics, laboratory parameters obtained from the initial blood samples collected at PICU admission, and the calculated pSOFA and PRISM IV scores were recorded.

The associations of RDW and MPV with CRP, lactate, procalcitonin, pSOFA score, and PRISM IV score were evaluated in critically ill patients diagnosed with sepsis or septic shock. Sepsis and septic shock were defined according to the criteria established by the International Pediatric Sepsis Consensus Conference. These criteria were selected because they were the sepsis/septic shock definitions used in the study institution during the study period and allowed consistent retrospective identification of eligible patients.

For PRISM IV calculation, laboratory variables obtained from the 2 hours preceding PICU admission through the first 4 hours after admission were included, whereas physiological variables were limited to those measured during the first 4 hours following PICU admission. The PRISM IV score incorporates patient age, source of admission (operating room, emergency department, hospital ward, or referral from another hospital), cardiopulmonary resuscitation within the 24 hours preceding PICU admission, presence of acute or chronic malignancy, primary dysfunction of low-risk organ systems (including endocrine, hematologic, musculoskeletal, and primary renal disorders), and the neurological/non-neurological diagnostic category.

The pSOFA score consists of six organ system components: respiratory ($\text{PaO}_2/\text{FiO}_2$ or $\text{SpO}_2/\text{FiO}_2$ ratio), hepatic (serum bilirubin concentration), hematologic (platelet count), cardiovascular (age-adjusted mean arterial pressure or the requirement for vasoactive agents such as dopamine, dobutamine, epinephrine, or norepinephrine), neurological (Glasgow Coma Scale score), and renal (age-adjusted serum creatinine level). Each organ system was assigned a score ranging from 0 to 4 according to the worst recorded value within the first 24 hours of PICU admission, and the total pSOFA score (range, 0-24) was calculated by summing the individual organ scores. Variables not measured during the first 24 hours were considered normal, consistent with the original pSOFA criteria.

Patients were followed until discharge from the PICU. The primary outcome was 28-day all-cause mortality. Secondary outcomes included a PICU length of stay of at least 48 hours and the total duration of PICU hospitalization.

Inclusion and Exclusion Criteria

Patients with hematologic or oncologic disorders (patients with primary non-malignant chronic hematologic disorders (e.g., hemoglobinopathies, hereditary anemias) or those who had received erythrocyte or platelet transfusions within the preceding three months were excluded, as

these factors independently alter baseline RDW and MPV parameters. Patients with hematologic malignancies (such as leukemias or lymphomas) presenting with acute sepsis were included, provided they had no recent transfusion history. Patients with cyanotic congenital heart disease, chronic liver disease, chronic kidney disease, or inherited metabolic disorders were excluded because these underlying conditions could independently influence RDW and MPV values. Patients who had received erythrocyte or platelet transfusions within the preceding three months for any reason, had a previous PICU admission with sepsis or septic shock during the study period, or had incomplete clinical or laboratory data were also excluded.

The admission-diagnosis categories were retained as recorded in the clinical dataset. The hematological admission-diagnosis category therefore refers to the coded reason for admission and does not alter the prespecified exclusion of patients with underlying hematologic or oncologic disorders.

Ethical Approval

The study protocol was approved by the Clinical Research Ethics Committee of the University of Health Sciences, İzmir Dr. Behçet Uz Pediatric Diseases and Surgery Training and Research Hospital (protocol no: 2017/122; decision no: 2017/01-03, date: 26.01.2017). The study was conducted in accordance with the ethical principles of the Declaration of Helsinki.

Statistical Analysis

Statistical analyses were performed using SPSS version 26.0 (IBM Corp., Armonk, NY, USA) and Python (pandas, SciPy, and scikit-learn libraries). The distribution of continuous variables was assessed using the Shapiro-Wilk test. Variables that did not follow a normal distribution were expressed as median (minimum-maximum). Comparisons between groups were performed using the Mann-Whitney U test. Associations between variables were evaluated using Spearman's rank correlation analysis.

The diagnostic performance of each biomarker for 28-day mortality was assessed using ROC curve analysis. The area under the ROC curve (AUC) and corresponding 95% confidence intervals were estimated using bootstrap resampling with 2,000 iterations. Optimal cut-off values were determined using the Youden index (J : sensitivity + specificity - 1). A two-sided p-value <0.05 was considered statistically significant.

Results

Demographic and Clinical Characteristics

A total of 243 pediatric patients were included in the study. The median age was 24 months (range, 1-216 months), and 146 patients (60.1%) were male, while 97 (39.9%) were female.

Respiratory diseases were the most common primary admission diagnosis (32.5%), followed by sepsis (21.8%) and neurological disorders (11.1%).

Vasoactive medications were administered to 60 patients (24.7%). Microbiological evaluation revealed no pathogen growth in 168 patients (69.1%). Among patients with positive cultures, blood cultures (11.9%) and urine cultures (10.3%) were the most frequently identified sources of microbiological positivity.

At the end of the 28-day follow-up period, 221 patients (90.9%) were discharged alive, whereas 22 patients (9.1%) died. The median length of PICU stay was 5 days (range, 1-136 days), and the median duration of mechanical ventilation was 0 days (range, 0-70 days).

At admission, the median serum albumin concentration was 3.7 g/dL (range, 1.0-5.1 g/dL), the median PRISM IV score was 3 (range, 0-39), and the median pSOFA score was 3 (range, 0-13). The median RDW was 14.4% (range, 12.4-17.4%), median MPV was 10.5 fL (range, 9.1-12.4 fL), median procalcitonin level was 3.22 ng/mL (range, 0.76-12.50 ng/mL), median CRP concentration was 46.3 mg/L (range, 3.0-249.6 mg/L), and median lactate level was 1.6 mmol/L (range, 0.5-7.3 mmol/L) (Table 1).

Comparison Between the Sepsis and Septic Shock Groups

Patients with septic shock exhibited significantly higher RDW, MPV, CRP, procalcitonin, and lactate levels, as well as significantly higher PRISM IV and pSOFA scores, compared with patients with sepsis (all $p < 0.001$) (Table 2).

Correlation Analyses

Spearman correlation analysis performed in the overall study population ($n=243$) demonstrated statistically significant positive correlations between RDW, MPV, CRP, procalcitonin, and lactate levels and both PRISM IV and pSOFA scores (all $p < 0.001$). The strongest correlation was observed between procalcitonin and the pSOFA score ($r=0.87$), whereas the weakest correlation was found between MPV and the PRISM IV score ($r=0.31$). RDW demonstrated a stronger correlation with clinical severity scores ($r=0.49$ - 0.53) than MPV ($r=0.31$ - 0.33) (Table 3).

ROC Curve Analyses for Predicting 28-day Mortality

Among the individual biomarkers evaluated for predicting 28-day mortality, lactate (AUC: 0.99) and CRP (AUC: 0.98) demonstrated the highest discriminative performance. These were followed by the pSOFA score (AUC: 0.89), procalcitonin (AUC: 0.87), RDW (AUC: 0.85), and the PRISM IV score (AUC: 0.82), whereas MPV showed the lowest discriminative ability (AUC: 0.73). A combined logistic regression model incorporating RDW, CRP, and MPV predicted mortality with excellent discriminatory performance (AUC: 0.99) (Table 4).

Table 1. Demographic, clinical, and laboratory characteristics of the patients included in the study

	n (%) / median (min-max)
Age (months)	24 (1-216)
Male, n (%)	146 (60.1%)
Female, n (%)	97 (39.9%)
Patients with underlying comorbidities	
Respiratory	79 (32.5%)
Sepsis	53 (21.8%)
Neurological	27 (11.1%)
Shock	17 (7.0%)
Metabolic	17 (7.0%)
Cardiac	15 (6.2%)
Hematological	8 (3.3%)
Other (trauma/postoperative)	4 (1.6%)
Patients with no underlying comorbidities	23 (0.09%)
Vasoactive drugs (yes)	60 (24.7%)
Positive microbiological culture	75 (30.9%)
Blood	29 (11.9%)
Urine	25 (10.3%)
Body fluid	9 (3.7%)
CSF	6 (2.5%)
Tracheal aspirate	6 (2.5%)
28-day mortality	22 (9.1%)
Length of PICU stay (days)	5 (1-136)
Duration of mechanical ventilation (days)	0 (0-70) (IQR)-(mean: 4.25 days)
Serum albumin (g/dL)	3.7 (1.0-5.1)
Note: Data are presented as n (%) or median (minimum-maximum) PICU: Pediatric intensive care unit, IQR: Interquartile range	

Age-based Subgroup ROC Analyses

In patients younger than 12 months, CRP (AUC: 1.00) and RDW (AUC: 0.81) demonstrated high discriminative performance, whereas MPV showed relatively poor predictive ability (AUC: 0.62). In contrast, among patients aged ≥ 12 months, both CRP (AUC: 0.98) and RDW (AUC: 0.87) maintained strong predictive performance, while the discriminative ability of MPV increased modestly compared with the infant subgroup (AUC: 0.78). In both age groups, the combined biomarker model (RDW+CRP+MPV) achieved comparable or superior discriminative performance compared with the individual biomarkers alone (Table 5).

Comparison of Admission Biomarkers and Severity Scores According to 28-day Mortality

When baseline parameters were compared between survivors and non-survivors at 28-days, non-survivors demonstrated significantly higher admission RDW, MPV, CRP, procalcitonin, and lactate levels, as well as higher

PRISM IV and pSOFA scores compared with survivors (all $p < 0.001$) (Table 6).

Discussion

In the present study, we investigated the associations of RDW, CRP, MPV, procalcitonin, and lactate with disease severity and mortality in 243 pediatric patients diagnosed

with sepsis or septic shock. Disease severity was evaluated using the PRISM IV and pSOFA scoring systems, ROC curve analysis, and age-based subgroup analyses. All evaluated biomarkers demonstrated significant positive correlations with both PRISM IV and pSOFA scores. Lactate and CRP exhibited the greatest discriminative ability for predicting 28-day mortality, whereas RDW proved to be a

Table 2. Comparison of the sepsis and septic shock groups

	Sepsis (n=183)	Septic shock (n=60)	p-value
RDW (%)	14.20 (12.40-16.60)	14.95 (13.40-17.40)	<0.001
MPV (fL)	10.50 (9.10-12.40)	10.80 (9.50-12.20)	<0.001
CRP (mg/L)	41.20 (3.00-140.00)	73.10 (14.70-249.60)	<0.001
Procalcitonin (ng/mL)	2.74 (0.76-9.12)	5.15 (1.35-12.50)	<0.001
Lactate (mmol/L)	1.50 (0.50-5.30)	2.30 (0.50-7.30)	<0.001
PRISM IV	2 (0-16)	6 (0-39)	<0.001
pSOFA	2 (0-9)	4.5 (0-13)	<0.001

Data are presented as median (minimum-maximum). Comparisons between groups were performed using the Mann-Whitney U test. RDW: Red cell distribution width, MPV: Mean platelet volume, CRP: C-reactive protein, PRISM IV: Pediatric Risk of Mortality IV, pSOFA: Pediatric Sequential Organ Failure Assessment

Table 3. Correlation of biomarkers with PRISM IV and pSOFA scores

	r-value	p-value
RDW - PRISM IV	0.49	<0.001
RDW - pSOFA	0.53	<0.001
MPV - PRISM IV	0.31	<0.001
MPV - pSOFA	0.33	<0.001
CRP - PRISM IV	0.73	<0.001
CRP - pSOFA	0.79	<0.001
Procalcitonin - PRISM IV	0.79	<0.001
Procalcitonin - pSOFA	0.87	<0.001
Lactate - PRISM IV	0.62	<0.001
Lactate - pSOFA	0.64	<0.001

Note: r-value = Spearman's correlation coefficient. PRISM IV: Pediatric Risk of Mortality IV, pSOFA: Pediatric Sequential Organ Failure Assessment, RDW: Red cell distribution width, MPV: Mean platelet volume, CRP: C-reactive protein

Table 4. Results of ROC analysis in predicting 28-day mortality

	AUC (95% CI)	Cut-off	Sensitivity (%)	Specificity (%)
CRP (mg/L)	0.98 (0.96-1.00)	≥ 88.5	95.5	92.8
RDW (%)	0.85 (0.75-0.92)	≥ 14.7	90.9	65.6
MPV (fL)	0.73 (0.62-0.82)	≥ 10.7	77.3	64.3
Procalcitonin (ng/mL)	0.87 (0.76-0.96)	≥ 6.05	72.7	92.8
Lactate (mmol/L)	0.99 (0.99-1.00)	≥ 3.20	100.0	97.3
PRISM IV	0.82 (0.69-0.93)	≥ 9	72.7	83.7
pSOFA	0.89 (0.79-0.97)	≥ 7	77.3	94.1
RDW+CRP+MPV (combined)	0.99 (0.97-1.00)	-	-	-

Optimal cut-off values were determined using the Youden index. ROC: Receiver operating characteristic, AUC: Area under the curve, CI: Confidence interval, CRP: C-reactive protein, MPV: Mean platelet volume, RDW: Red cell distribution width, PRISM IV: Pediatric Risk of Mortality IV, pSOFA: Pediatric Sequential Organ Failure Assessment

Table 5. ROC analysis of biomarkers for predicting 28-day mortality according to age groups

Age groups		AUC (95% CI)	Cut-off	Sensitivity (%)	Specitifiity (%)
<12 months (n=77)	RDW (%)	0.81 (0.63-0.95)	≥15.2	72.7	87.9
	CRP (mg/L)	1.00 (0.98-1.00)	≥98.7	100.0	97.0
	MPV (fL)	0.62 (0.45-0.78)	≥10.6	81.8	43.9
	Combined (RDW+CRP+MPV)	1.00 (1.00-1.00)	-	-	-
≥12 months (n=166)	RDW (%)	0.87 (0.79-0.95)	≥14.7	100.0	67.7
	CRP (mg/L)	0.98 (0.93-1.00)	≥88.5	90.9	92.9
	MPV (fL)	0.78 (0.60-0.91)	≥10.8	81.8	74.2
	Combined (RDW+CRP+MPV)	0.99 (0.96-1.00)	-	-	-

ROC: Receiver operating characteristic, AUC: Area under the curve, CI: Confidence interval, CRP: C-reactive protein, MPV: Mean platelet volume, RDW: Red cell distribution width

Table 6. Comparison of admission biomarkers and severity scores according to 28-day mortality

Variable	Survivors (n=221), median (min-max)	Deaths (n=22), median (min-max)	p-value
RDW (%)	14.30 (12.40-16.70)	15.50 (13.60-17.40)	<0.001
MPV (fL)	10.50 (9.10-12.00)	10.80 (10.00-12.40)	<0.001
CRP (mg/L)	43.90 (3.00-123.50)	139.70 (67.90-249.60)	<0.001
Procalcitonin (ng/mL)	3.01 (0.76-7.67)	7.18 (1.96-12.50)	<0.001
Lactate (mmol/L)	1.60 (0.50-5.20)	4.90 (3.20-7.30)	<0.001
PRISM IV	3.00 (0.00-21.00)	12.50 (0.00-39.00)	<0.001
pSOFA	2.00 (0.00-9.00)	8.00 (1.00-13.00)	<0.001

Continuous variables were compared between the sepsis and septic shock groups using the Mann-Whitney U test and are presented as median (interquartile range) (p<0.01)
RDW: Red cell distribution width, MPV: Mean platelet volume, CRP: C-reactive protein, PRISM IV: Pediatric Risk of Mortality IV, pSOFA: Pediatric Sequential Organ Failure Assessment

more consistent and robust prognostic marker than MPV. Furthermore, combining these biomarkers significantly improved diagnostic performance, and their predictive value varied according to patient age.

Sepsis and septic shock remain major causes of morbidity and mortality in both pediatric and adult populations, particularly when diagnosis and treatment are delayed. Consequently, there is a growing need for readily available, inexpensive biomarkers capable of identifying disease severity at an early stage (6). Although prognostic scoring systems such as PRISM IV and pSOFA are widely used in clinical practice, their calculation requires numerous clinical and laboratory variables and may therefore be time-consuming. In this context, evaluating the prognostic value of parameters routinely obtained from complete blood count analysis and acute-phase reactants is of considerable clinical importance.

In our cohort, lactate and CRP demonstrated the highest AUC values among the individual biomarkers for predicting 28-day mortality. Notably, the combined model incorporating RDW, CRP, and MPV achieved a discriminative performance comparable to that of lactate alone. As lactate is a direct indicator of tissue hypoperfusion and cellular metabolic

stress, its strong association with mortality during the early stages of septic shock is biologically plausible. However, lactate measurement may not be readily available in all healthcare settings, particularly in secondary-care hospitals. Our findings suggest that routinely available laboratory parameters may provide comparable prognostic information, thereby facilitating early risk assessment in resource-limited settings.

Previous pediatric studies have demonstrated that CRP reflects the severity of systemic inflammation and is associated with mortality in patients with sepsis and septic shock (7). The significantly elevated CRP levels observed in patients with septic shock in our study further support its role as a marker of systemic inflammatory response, consistent with previous reviews evaluating biomarkers in sepsis (8).

Procalcitonin is one of the most extensively investigated biomarkers for the diagnosis and risk stratification of bacterial infections and sepsis (9). Numerous studies have demonstrated that elevated procalcitonin concentrations correlate with the presence of systemic infection, greater disease severity, and increased mortality risk (10). Furthermore, elevated procalcitonin levels at clinical

presentation have been associated with intensive care admission and adverse clinical outcomes across various infectious diseases, particularly community-acquired pneumonia (11). Consistent with these reports, procalcitonin levels were significantly elevated in patients with septic shock compared to those with sepsis in our cohort, and ROC curve analysis demonstrated strong discriminative performance for 28-day mortality. While these findings highlight the utility of procalcitonin as a marker of disease severity and short-term prognosis in critically ill children, prospective case-control studies are required to directly evaluate and compare its baseline diagnostic sensitivity relative to CRP and hematological indices.

RDW has recently emerged as an indirect marker of inflammation, oxidative stress, and bone marrow dysfunction (12). In our study, RDW demonstrated significant positive correlations with both PRISM IV and pSOFA scores and showed superior predictive performance for mortality compared with MPV (AUC: 0.85 vs. 0.73). These findings suggest that RDW may represent a more reliable prognostic biomarker than MPV in pediatric sepsis. Although several studies have reported an association between RDW and mortality in adult patients with sepsis (13,14), evidence in pediatric populations remains limited. Our findings, therefore, provide additional support for the use of RDW as an inexpensive, readily available biomarker reflecting disease severity and mortality in pediatric sepsis and septic shock.

MPV is considered a marker of platelet activation and systemic inflammatory response and is therefore biologically relevant in the pathophysiology of sepsis, in which platelets contribute to inflammation, microvascular thrombosis, and endothelial injury (15). In the present study, MPV demonstrated statistically significant but relatively weak correlations with PRISM IV and pSOFA scores ($r=0.31-0.33$) and showed the lowest discriminative ability for predicting mortality (AUC: 0.73). These findings indicate that the prognostic value of MPV alone may be limited. The dynamic nature of platelet activation and peripheral platelet consumption during sepsis may explain the greater biological variability of MPV compared with RDW and CRP. Accordingly, MPV appears to be more informative when interpreted in combination with other inflammatory biomarkers rather than as an isolated parameter.

Age-stratified subgroup analyses provided additional exploratory insights in our study. Among patients younger than 12 months, CRP demonstrated high discriminative performance, which may reflect the relatively pronounced acute-phase response during infancy. Conversely, the lower performance of MPV in this age group might be related to the immaturity of the hematopoietic system and age-dependent differences in platelet kinetics. In children aged ≥ 12 months, the discriminative performance of RDW and MPV appeared to improve, potentially aligning

with the progressive maturation of erythropoietic and thrombopoietic functions with age. However, given the limited number of outcome events within these age subgroups, these findings—and the potential utility of age-specific cut-off values—should be interpreted with caution and regarded primarily as hypothesis-generating, warranting further validation in larger cohorts (16).

The excellent AUC values obtained from the combined RDW, CRP, and MPV model indicate that the simultaneous assessment of these biomarkers provides superior predictive performance compared with their individual use. Similar findings have been reported in previous studies investigating multimarker approaches (17,18). From a practical perspective, this finding suggests that disease severity and mortality may be predicted with high accuracy using only routinely available laboratory parameters, thereby facilitating early clinical risk stratification without requiring complex or costly investigations.

Study Limitations

Several limitations of this study should be acknowledged. First, the retrospective, single-center design may limit the generalizability of our findings to broader pediatric populations. Second, because only laboratory values obtained at PICU admission were analyzed, dynamic changes and temporal trends in biomarker levels during the clinical course could not be evaluated. Third, as the primary objective of this study was early risk stratification among critically ill children with sepsis requiring intensive care, an independent control group consisting of healthy children or patients with non-septic mild infections was not included. Consequently, the diagnostic specificity of the evaluated biomarkers could not be fully assessed. Finally, the combined model analysis in patients younger than 12 months and the ROC analyses for disease were exploratory in nature because of the limited number of outcome events. In this study, the “International Pediatric Sepsis Consensus Conference” criteria were used for the diagnosis of sepsis and septic shock. However, while the study was ongoing, a new study was also planned using the new “Phoenix” diagnostic criteria published in 2024. Nevertheless, we believe that our study makes a meaningful contribution to the existing literature by demonstrating strong associations between inexpensive, routinely available biomarkers and validated clinical severity scores in a homogeneous pediatric intensive care population.

Future Research

In light of the present findings and the limitations identified, future investigations should focus on larger, multicenter prospective cohort studies to validate these results. New studies should be conducted using the Phoenix diagnostic criteria published in 2024. Serial assessment of biomarker kinetics throughout the PICU stay may provide a more dynamic approach to predicting treatment response and

clinical outcomes. In addition, studies incorporating well-defined control groups are needed to further establish the diagnostic specificity of these biomarkers. Finally, combined biomarker models and age-specific cut-off values, particularly for vulnerable populations such as infants, should be externally validated in independent pediatric cohorts before being incorporated into routine clinical practice.

Conclusion

This study demonstrates that RDW, CRP, MPV, procalcitonin, and lactate are readily available and cost-effective biomarkers associated with disease severity and 28-day mortality in pediatric sepsis and septic shock. Among the individual biomarkers evaluated, procalcitonin, lactate, and CRP exhibited the strongest discriminative performance, whereas RDW proved to be a more consistent prognostic marker than MPV. Moreover, the combined assessment of RDW, CRP, and MPV provided superior discriminative performance compared with the use of each biomarker individually.

From a clinical perspective, the combined application of these biomarkers using age-specific cut-off values may facilitate early risk stratification in pediatric patients with sepsis and septic shock. Nevertheless, these findings should be confirmed by large-scale, prospective, multicenter studies, particularly those including a greater number of patients with disease, to further validate the clinical utility and generalizability of these biomarkers. This study does not establish diagnostic accuracy for sepsis or the value of serial biomarker monitoring, and the reported prognostic findings require validation in larger, prospective, multicenter cohorts.

Ethics

Ethics Committee Approval: The study protocol was approved by the Clinical Research Ethics Committee of the University of Health Sciences, Dr. Behçet Uz Pediatric Diseases and Surgery Training and Research Hospital (protocol no. 2017/122; decision no. 2017/01-03, date: 26.01.2017).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: F.S., Ö.S., G.A., E.S., Concept: F.S., H.A., Design: F.S., H.A., Data Collection or Processing: F.S., Analysis or Interpretation: G.C., G.A., Literature Search: G.C., Ö.S., G.A., Writing: F.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. Recher M, Leteurre S, Canon V, Baudalet JB, Lockhart M, Hubert H. Severity of illness and organ dysfunction scoring systems in pediatric critical care: the impacts on clinician's practices and the future. *Front Pediatr*. 2022;10:1054452.
2. Arı HF, Reşitoğlu S, Tuncel MA, Şerbetçi MC. Comparison between mortality scoring systems in pediatric intensive care unit reliability and effectiveness. *Pam Med J*. 2024;17:664-73.
3. Ala-Selek S, Akalın H, Azap A, Kurt-Azap Ö, Demirkıran O, Doğan D, et al. [Use of procalcitonin and C-reactive protein in the diagnosis and treatment of infectious diseases: National Consensus Report]. *Klinik Derg*. 2025;38:62-96. Turkish.
4. Aker Karagöz Y, Şahin İ, Karagöz F, Şit D. Evaluation of relationship between uric acid and CRP, RDW, and MPV as cardiovascular risk factors in patients with hypertension. *Bezmialem Science*. 2017;5:16-21.
5. Mishra N, Rana A, Kumar P, Gupta R, Kumar M. Complete blood count (CBC) parameters as a cost-effective tool for early diagnosis of pediatric sepsis: a retrospective cross-sectional study. *Cureus*. 2025;17:e89661.
6. Weiss SL, Peters MJ, Oczkowski SJW, Belley-Cote E, Buysse C, Choong KLM, et al. Surviving sepsis campaign international guidelines for the management of sepsis and septic shock in children 2026. *Intensive Care Med*. 2026;52:937-83.
7. Usluer E, Anıl AB, Anıl M, Kamit F, Altuğ Ü, Özçifçi G, et al. Clinical outcomes and mortality predictors in patients hospitalized in the pediatric intensive care unit due to sepsis. *J Dr Behçet Uz Child Hosp*. 2025;15:111-20.
8. Tyagi N, Gawhale S, Patil MG, Tambolkar S, Salunkhe S, Mane SV. Comparative analysis of C-reactive protein and procalcitonin as biomarkers for prognostic assessment in pediatric sepsis. *Cureus*. 2024;16:e65427.
9. Wong HR. Pediatric sepsis biomarkers for prognostic and predictive enrichment. *Pediatr Res*. 2022;91:283-8.
10. Menon K, Schlapbach LJ, Akech S, Argent A, Biban P, Carrol ED, et al.; Pediatric Sepsis Definition Taskforce of the Society of Critical Care Medicine. Criteria for pediatric sepsis-a systematic review and meta-analysis by the pediatric sepsis definition taskforce. *Crit Care Med*. 2022;50:21-36.
11. Lucas G, Bartolf A, Kroll N, De Thabrew AU, Murtaza Z, Kumar S, et al. Procalcitonin (PCT) level in the emergency department identifies a high-risk cohort for all patients treated for possible sepsis. *EJIFCC*. 2021;32:20-6.
12. Wang F, Pan W, Pan S, Ge J, Wang S, Chen M. Red cell distribution width as a novel predictor of mortality in ICU patients. *Ann Med*. 2011;43:40-6.
13. Aygün F. The red cell distribution width can be associated with sepsis in patients older than two-year-old. *Kocatepe Med J*. 2020;21:35-42.
14. Yu Y, Hu X, Shen Y. Red blood cell distribution width as a predictor of mortality in critically ill pediatric patients: a systematic review and meta-analysis. *Front Pediatr*. 2025;13:1646179.
15. Sunkonkit K, Chai-Adisaksopha C, Natesirinilkul R, Phinyo P, Trongtrakul K. Admission red blood cell distribution width and mean platelet volume as predictors of mortality in

- the pediatric intensive care unit: a five-year single-center retrospective study. *J Clin Med*. 2025;14:3839.
16. Diyeke E. Does C-reactive protein, mean platelet volume and red cell distribution width predict ventilator-associated events in mechanically ventilated patients? *J Immunol Clin Microbiol*. 2020;5:73-81.
17. Yeşildağ M, Özkan Bakdik B, Balasar B, Eroğlu E. Evaluation of the relationship between biomarkers and disease severity in patients with community-acquired pneumonia. *Eur J Ther*. 2024;30:354-61.
18. Zhang YL, Yuan G, Xiao YY, Li XJ. Prognostic value of combined RDW, NLR, PLR, and mpv detection in patients with acute exacerbation of chronic obstructive pulmonary disease: a single-center retrospective cohort study. *BMC Pulm Med*. 2026;26:158.



Radiologically Presumed Needle-tract Seeding After Radiofrequency Ablation for Hepatocellular Carcinoma: A Case Report

Hepatoselüler Karsinomda Radyofrekans Ablasyon Sonrası Radyolojik Olarak Düşünülen İğne Traktı Tümör Ekimi: Olgu Sunumu

¹ Hacer KUNDAKCIOĞLU¹, ² Hüseyin TOPRAK², ³ Adem AKÇAKAYA³

¹Afşin State Hospital, Clinic of General Surgery, Kahramanmaraş, Türkiye

²Bezmialem Vakıf University Faculty of Medicine, Department of Radiology, İstanbul, Türkiye

³Bezmialem Vakıf University Faculty of Medicine, Department of General Surgery, İstanbul, Türkiye

ABSTRACT

Radiofrequency ablation (RFA) is an effective and widely used percutaneous treatment for hepatocellular carcinoma (HCC), particularly in patients awaiting liver transplantation. However, despite technological advances, tumor seeding along the needle tract remains a rare but clinically significant complication. A 59-year-old male patient with Child-Pugh A cirrhosis secondary to hepatitis B virus/hepatitis D virus infection and a prior diagnosis of HCC was found to have a newly developed 18×12 mm subcapsular lesion in segment VI of the liver on follow-up magnetic resonance imaging (MRI). RFA was performed prior biopsy. Three months later, MRI demonstrated newly developed nodular lesions along the ablation tract surrounding the ablation zone, highly suggestive of needle-tract seeding. The patient received systemic therapy including lenvatinib, sorafenib, and regorafenib, as well as repeated RFA and transarterial chemoembolization procedures. However, liver transplantation could not be performed because a suitable donor was unavailable, and the patient remains under ongoing oncologic follow-up. Although the overall incidence of neoplastic seeding after RFA is low (0.6-1.4%), several factors—including subcapsular tumor location, poor differentiation, large tumor size, elevated alpha-fetoprotein levels, and pre-ablation biopsy—have been associated with increased risk. Needle tract ablation and

ÖZ

Radyofrekans ablasyon (RFA), özellikle karaciğer nakli bekleyen hastalarda hepatoselüler karsinomun (HCC) tedavisinde etkili ve yaygın olarak kullanılan perkütan bir yöntemdir. Ancak, teknolojiye ileriye gidenler rağmen, iğne traktı boyunca tümör ekilmesi nadir görülen fakat klinik olarak önemli bir komplikasyon olmaya devam etmektedir. Hepatit B virüsü/hepatit D virüsü ilişkili Child-Pugh A sirozu olan, daha önce HCC tanısı almış 59 yaşındaki erkek hastada takip manyetik rezonans görüntülemesinde (MRG) karaciğerin VI. segmentinde, subkapsüler yerleşimli 18×12 mm boyutunda yeni gelişen bir lezyon saptanmıştır. Biyopsi yapılmaksızın RFA uygulanmıştır. Üç ay sonra yapılan MRG'de ablatif zon çevresinde iğne traktı boyunca yeni gelişen ve iğne traktı ekimini düşündüren nodüler lezyonlar izlenmiştir. Hasta lenvatinib, sorafenib ve regorafenib içeren sistemik tedaviler ile tekrar RFA ve transarteriyel kemoembolizasyon uygulamaları almıştır; ancak uygun verici bulunamadığı için karaciğer transplantasyonu gerçekleştirilememiştir. Hastanın onkolojik takibi RFA sonrası neoplastik ekilme oranı genel olarak düşük (%0,6-1,4) olmakla birlikte; subkapsüler yerleşim, kötü diferansiasyon, büyük tümör boyutu, yüksek alfa-fetoprotein düzeyi ve ablasyon öncesi biyopsi gibi faktörlerin artmış risk ile ilişkili olduğu bildirilmiştir. İğne traktı ablasyonu ve dikkatli hasta seçiminin bu riski azaltabildiği gösterilmiştir. Ayrıca indosiyanin yeşili

Address for Correspondence: Hacer Kundakcıoğlu MD, Afşin State Hospital, Clinic of General Surgery, Kahramanmaraş, Türkiye

E-mail: haker.kundakcioglu@hotmail.com

ORCID IDs of the authors: H.K.: 0000-0002-1064-2483, H.T.: 0000-0002-6641-7648, A.A.: 0000-0003-3116-7033

Cite this article as: Kundakcıoğlu H, Toprak H, Akçakaya A. Radiologically presumed needle-tract seeding after radiofrequency ablation for hepatocellular carcinoma: a case report. Bezmialem Science. 2026;14(4):410-414



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

Received: 27.12.2022

Accepted: 11.11.2025

Epub: 03.07.2026

Published date: 26.10.2026

ABSTRACT

Careful patient selection have been shown to reduce this risk. Additionally, novel imaging modalities such as indocyanine green near-infrared fluorescence may facilitate intraoperative detection and complete excision of seeded foci. This case suggests that tumor seeding may occur after RFA even in the absence of pre-procedural biopsy. In patients with subcapsular lesions, meticulous procedural technique, careful evaluation of risk factors, and close post-procedural surveillance are essential for optimizing clinical outcomes.

Keywords: Tumor seeding, hepatocellular carcinoma, radiofrequency ablation

ÖZ

yakın kızılötesi floresan görüntüleme gibi yeni görüntüleme yöntemleri, ekilmiş odakların intraoperatif tespitine ve tam rezeksiyonuna olanak sağlamaktadır. Bu olgu, biyopsi yapılmamış olgularda dahi RFA sonrası iğne traktı ekiminin gelişebileceğini düşündürmektedir. Özellikle subkapsüler lezyonlarda işlem tekniğine dikkat edilmesi, risk faktörlerinin değerlendirilmesi ve yakın takip, klinik sonuçların iyileştirilmesi açısından büyük öneme sahiptir.

Anahtar Kelimeler: Tümör ekimi, hepatosellüler karsinom, radyofrekans ablasyon

Introduction

The use of percutaneous ablative techniques for the treatment of liver tumors has increased substantially over the past decades. Percutaneous ethanol injection (PEI), radiofrequency ablation (RFA), and microwave ablation (MWA) are viable curative techniques for hepatocellular carcinoma (HCC) (1).

Studies have shown that percutaneous treatment of HCC is a cost-effective bridging strategy in patients whose expected waiting time for liver transplantation exceeds six months (2).

However, when tumors are located in specific anatomical regions, such as beneath Glisson's capsule, in an exophytic position, or adjacent to critical structures, percutaneous treatment may be associated with complications including incomplete ablation, hemorrhage, injury to adjacent organs, and needle-tract seeding (3). Subcapsular location, poor tumor differentiation, tumor size greater than 3 cm, elevated serum alpha-fetoprotein (AFP) levels, and diagnostic biopsy before ablation have been reported as risk factors for needle-tract seeding (4-6).

Case Report

A 59-year-old male patient had been under follow-up since 2002 for HBV, HDV, and chronic liver disease. The patient progressed to Child-Pugh class A cirrhosis, and he was diagnosed with HCC 6.5 years ago. The patient had previously undergone transarterial chemoembolization and systemic therapy for HCC. He was subsequently followed with serial magnetic resonance imaging (MRI) examinations. His laboratory tests were carcinoembryonic antigen, 10 ng/mL; carbohydrate-related antigen 19-9, 28 U/mL; and AFP, 2411 ng/mL.

During follow-up MRI, a subcapsular lesion measuring approximately 18×12 mm was detected in segment VI of the right hepatic lobe. The lesion was considered suspicious for a new HCC lesion in a patient with a known history of HCC and markedly elevated AFP levels. Positron emission

tomography/computed tomography was performed as part of the pre-transplant evaluation and to further characterize the lesion. Because no significant FDG uptake was observed, histopathological confirmation or further imaging studies were recommended, considering that well-differentiated HCC may demonstrate low FDG avidity. Despite the recommendation for further diagnostic evaluation, RFA was subsequently performed for the segment VI lesion at another medical center. No biopsy was performed either before or after the ablative procedures.

Three months after the RFA, MRI demonstrated a 27×22 mm ablation zone in segment VI, accompanied by changes consistent with post-procedural hemorrhage (Figure 1). No significant residual or recurrent tumor was detected. One month later, newly developed nodular lesions measuring 10 mm along the capsular surface and 11 mm within the adjacent superior adipose tissue were detected near the RFA zone. Based on their anatomical continuity with the ablation tract and interval growth on serial MRI examinations, these lesions were considered radiologically suspicious for needle-tract seeding (Figure 2). A significant increase in lesion size was observed on subsequent MRI examinations.

The patient was subsequently referred to the oncology department, and systemic treatment with lenvatinib with pembrolizumab was initiated. As the lesion progressed, sorafenib was started, then switched to regorafenib. Subsequently, repeat RFA and transarterial chemoembolization were performed, resulting in partial regression of the lesions. Following these interventions, the AFP level decreased to 1400 ng/mL, and treatment with regorafenib was continued. Liver transplantation could not be performed because a suitable donor was not available. The patient remains under ongoing clinical and radiological follow-up.

Written informed consent was obtained from the patient for the publication of this case report and any accompanying images.

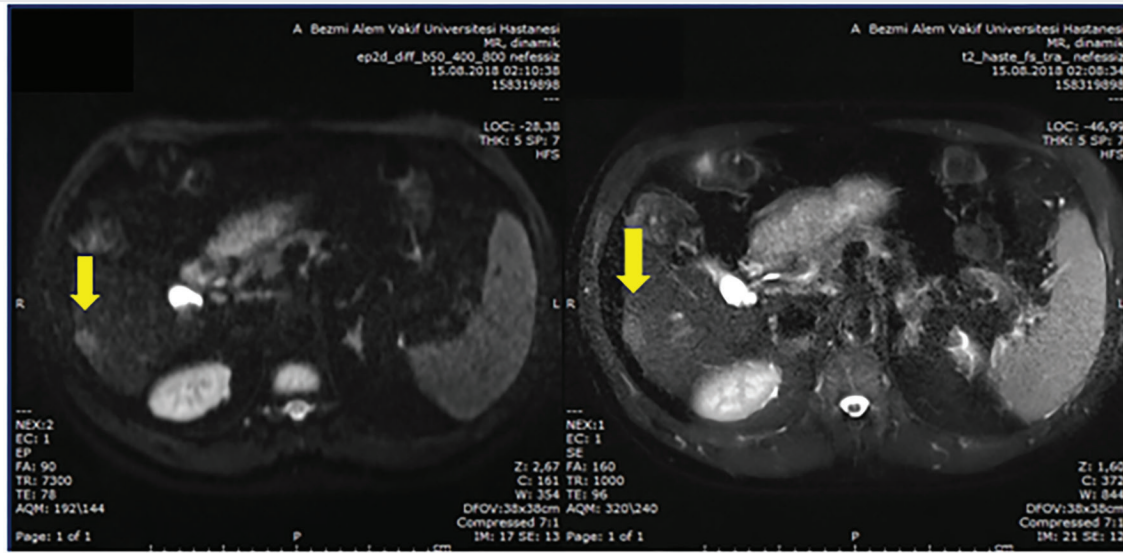


Figure 1. Axial magnetic resonance imaging demonstrating a 27×22 mm lesion located peripherally in segment VI of the right hepatic lobe, consistent with the post-ablation zone following radiofrequency ablation

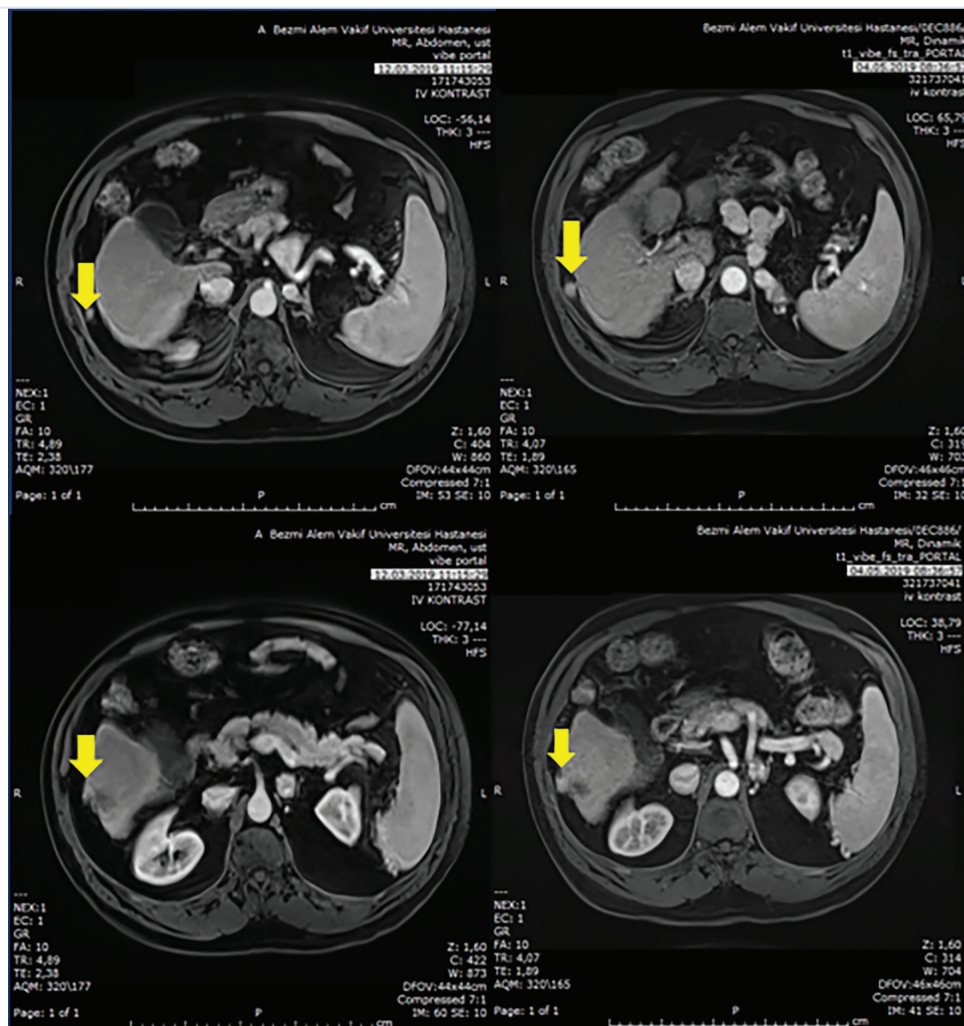


Figure 2. Follow-up magnetic resonance imaging showing newly developed nodular lesions adjacent to the radiofrequency ablation zone: a 10 mm nodule on the capsular surface and an 11 mm lesion within the adjacent adipose tissue, suggestive of neoplastic tract seeding

Discussion

Needle-tract seeding is a recognized complication of biopsy and percutaneous ablative techniques for HCC, including RFA, PEI, and MWA. Although reported rates vary among studies, the overall incidence remains low.

Llovet et al. (3) reported a tumor seeding rate of 12.5% after RFA for HCC, which is considerably higher than that reported in subsequent studies. Takamori et al. (7) reported a seeding rate of 5.1% in patients who underwent diagnostic biopsy before RFA. The overall incidence of tumor seeding after liver biopsy for HCC was reported to be approximately 2.7% (8). Moreover, in a multicenter study involving 1,314 patients, Livraghi et al. (1) reported similar seeding rates among the three participating centers (0.7%, 0.9%, and 1.4%). In another study involving 60 cirrhotic patients with subcapsular HCC treated with RFA or MWA, needle-tract seeding occurred in only one patient, who had undergone biopsy before ablation (9). According to the systematic review by Stigliano et al. (10), reported seeding rates were 2.29% after biopsy alone, 1.4% after PEI combined with biopsy, 0.61% after RFA without biopsy, 0.95% after RFA with biopsy, and 0.72% for biopsied and subsequently ablated liver nodules, including non-HCC lesions.

Studies support that RFA is an effective and safe bridging therapy until transplantation in patients with HCC (2). However, because needle-tract seeding may adversely affect clinical outcomes, careful patient selection and the implementation of preventive measures are warranted (11).

It has been shown in an experimental animal model that needle tract ablation in HCC reduces the seeding rate (12). In another study, when RFA was performed on subcapsular HCC, no tract seeding was observed when needle tract ablation was performed on tumors 4 cm and smaller (13). Intraoperative use of indocyanine green near-infrared fluorescence imaging has enabled the detection of seeded tissues and peritoneal implants along the RFA needle. This technique enables complete surgical excision of all tumor-positive areas, along with histopathological confirmation of HCC in the resected specimens (14).

The present case should be interpreted cautiously because histopathological confirmation of either the treated segment VI lesion or the subsequently developed nodules was unavailable. Nevertheless, the appearance of new nodules along the presumed needle trajectory shortly after RFA and their progressive enlargement on serial imaging strongly suggested radiologically presumed needle-tract seeding. Therefore, the diagnosis in this case remains presumptive and based primarily on imaging findings and temporal association with the intervention.

Although histopathological confirmation was unavailable, alternative diagnoses such as post-ablation inflammatory nodules or hemorrhagic changes were considered less likely because the lesions developed along the presumed needle trajectory and demonstrated progressive enlargement on serial MRI examinations.

Conclusion

Although no diagnostic biopsy was performed, radiologically presumed needle-tract seeding developed in our patient, suggesting that subcapsular tumor location and markedly elevated AFP levels may have contributed to the development of this complication.

Whenever possible, the diagnosis of HCC should rely on established radiological and clinical criteria to avoid unnecessary biopsy and its potential complications. These findings highlight the importance of meticulous procedural technique and careful patient selection in minimizing the risk of tumor dissemination after percutaneous interventions for HCC. The use of adjunctive technologies, such as tract ablation and intraoperative fluorescence imaging, may improve safety and oncologic outcomes in this patient group.

Ethics

Informed Consent: Written informed consent was obtained from the patient for the publication of this case report and any accompanying images.

Acknowledgements

The authors used artificial intelligence tools exclusively for English language editing and grammatical revision of the manuscript. No AI tools were used for data analysis, interpretation of results, or generation of scientific content.

Footnotes

Authorship Contributions

Surgical and Medical Practices: H.K., H.T., A.A., Concept: H.K., H.T., A.A., Design: H.K., H.T., A.A., Data Collection or Processing: H.K., Analysis or Interpretation: H.K., A.A., Literature Search: H.K., Writing: H.K.

Conflict of Interest: One of the authors of this article (A.A.) is a member of the Editorial Board of this journal. He had no involvement in the peer-review process or editorial decision regarding this manuscript. The peer-review process and editorial decision were handled independently by another editor and no conflict of interest was declared by the other authors.

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

References

1. Livraghi T, Lazzaroni S, Meloni F, Solbiati L. Risk of tumour seeding after percutaneous radiofrequency ablation for hepatocellular carcinoma. *Br J Surg*. 2005;92:856-8.
2. Fontana RJ, Hamidullah H, Nghiem H, Greenson JK, Hussain H, Marrero J, et al. Percutaneous radiofrequency thermal ablation of hepatocellular carcinoma: a safe and effective bridge to liver transplantation. *Liver Transpl*. 2002;8:1165-74.
3. Llovet JM, Vilana R, Brú C, Bianchi L, Salmeron JM, Boix L, et al.; Barcelona Clinic Liver Cancer (BCLC) Group. Increased risk of tumor seeding after percutaneous radiofrequency ablation for single hepatocellular carcinoma. *Hepatology*. 2001;33:1124-9.
4. Imamura J, Tateishi R, Shiina S, Goto E, Sato T, Ohki T, et al. Neoplastic seeding after radiofrequency ablation for hepatocellular carcinoma. *Am J Gastroenterol*. 2008;103:3057-62.
5. Perkins JD. Seeding risk following percutaneous approach to hepatocellular carcinoma. *Liver Transpl*. 2007;13:1603.
6. Zhong-Yi Z, Wei Y, Kun Y, Ying D, Wei W, Jung-Chieh L, et al. Needle track seeding after percutaneous radiofrequency ablation of hepatocellular carcinoma: 14-year experience at a single centre. *Int J Hyperthermia*. 2017;33:454-8.
7. Takamori R, Wong LL, Dang C, Wong L. Needle-tract implantation from hepatocellular cancer: is needle biopsy of the liver always necessary? *Liver Transpl*. 2000;6:67-72.
8. Silva MA, Hegab B, Hyde C, Guo B, Buckels JA, Mirza DF. Needle track seeding following biopsy of liver lesions in the diagnosis of hepatocellular cancer: a systematic review and meta-analysis. *Gut*. 2008;57:1592-6.
9. Francica G, Meloni MF, de Sio I, Smolock AR, Brace CL, Iadevaia MD, et al. Radiofrequency and microwave ablation of subcapsular hepatocellular carcinoma accessed by direct puncture: safety and efficacy. *Eur J Radiol*. 2016;85:739-43.
10. Stigliano R, Marelli L, Yu D, Davies N, Patch D, Burroughs AK. Seeding following percutaneous diagnostic and therapeutic approaches for hepatocellular carcinoma. What is the risk and the outcome? Seeding risk for percutaneous approach of HCC. *Cancer Treat Rev*. 2007;33:437-47.
11. Shirai K, Tamai H, Shingaki N, Mori Y, Moribata K, Enomoto S, et al. Clinical features and risk factors of extrahepatic seeding after percutaneous radiofrequency ablation for hepatocellular carcinoma. *Hepatol Res*. 2011;41:738-45.
12. Ham SJ, Choi Y, Kim KW, Cho YC, Woo DC, Lee CW, et al. Tract ablation after radiofrequency ablation to prevent viable tumor cell adhesion to the needle electrode. *Int J Hyperthermia*. 2020;37:1287-92.
13. Filippousis P, Sotiropoulou E, Manataki A, Konstantinopoulos O, Thanos L. Radiofrequency ablation of subcapsular hepatocellular carcinoma: single center experience. *Eur J Radiol*. 2011;77:299-304.
14. Nakamura M, Hayami S, Ueno M, Kawai M, Miyamoto A, Suzaki N, et al. Detection of needle tract implantation and peritoneal seeding after radiofrequency ablation using intraoperative near-infrared fluorescence system for recurrent hepatocellular carcinoma: a case report. *Surg Case Rep*. 2018;4:76.



Physiotherapy for Pelvic Dysfunction: A Comprehensive Review

Pelvik Disfonksiyonda Fizyoterapi: Kapsamlı Bir İnceleme

Özge BAYKAN ÇOPUROĞLU¹, Betül ERGÜN²

¹Kayseri University İncesu Ayşe and Saffet Arslan Health Services Vocational School, Department of Treatment and Rehabilitation, Physiotherapy Program, Kayseri, Türkiye

²Firat University Faculty of Health Sciences, Department of Physiotherapy and Rehabilitation, Elazığ, Türkiye

ABSTRACT

Pelvic floor dysfunction (PFD) includes a range of disorders such as urinary and fecal incontinence, pelvic organ prolapse, pelvic pain, and sexual dysfunction. These conditions affect millions globally, often leading to decreased quality of life. Physiotherapy has emerged as a first-line, non-invasive treatment with growing clinical support. To synthesize current evidence (2014-2024) regarding physiotherapy-based interventions for the treatment and prevention of PFD in general and special populations. A narrative literature review was conducted using peer-reviewed articles, randomized controlled trials, systematic reviews, and clinical guidelines. Interventions assessed include pelvic floor muscle training (PFMT), biofeedback, electrical stimulation, manual therapy, behavioral modification, and telerehabilitation. PFMT remains the most effective and widely recommended intervention, particularly for stress urinary incontinence and mild pelvic organ prolapse. Biofeedback enhances motor control and adherence, especially in older adults or postpartum women. Electrical stimulation benefits patients with poor voluntary contraction capacity. Manual therapy is effective in cases of pelvic pain associated with hypertonic muscle activity. Behavioral strategies, including bladder training and toileting education, further optimize outcomes. Emerging evidence supports the use of app-based PFMT and digital tools, though engagement varies. Physiotherapy shows positive outcomes in special populations including postpartum, postmenopausal, and post-prostatectomy patients. Physiotherapy offers an evidence-based, individualized approach for managing PFD across a range of populations. Continued research, broader

ÖZ

Pelvik taban disfonksiyonu (PTD), idrar ve gaita inkontinansı, pelvik organ sarkması, pelvik ağrı ve cinsel işlev bozukluğu gibi çeşitli rahatsızlıkları kapsar. Bu durumlar dünya genelinde milyonlarca kişiyi etkilemekte ve sıklıkla yaşam kalitesinde azalmaya yol açmaktadır. Fizyoterapi, giderek artan klinik kanıtlarla desteklenen, invaziv olmayan bir birinci basamak tedavi seçeneği olarak öne çıkmaktadır. Bu derleme çalışmasının amacı, 2014-2024 yılları arasında yayımlanan güncel literatürü temel alarak, genel ve özel popülasyonlarda PTD'nin tedavi ve önlenmesinde kullanılan fizyoterapi temelli yaklaşımları ortaya koymaktır. Bu amaçla anlatı türünde bir literatür taraması gerçekleştirilmiş; hakemli makaleler, randomize kontrollü çalışmalar, sistematik derlemeler ve klinik kılavuzlar incelenmiştir. Değerlendirilen müdahaleler arasında pelvik taban kas eğitimi (PFME), biyogeribildirim, elektrik stimülasyonu, manuel terapi, davranışsal değişim yaklaşımları ve telerehabilitasyon yer almaktadır. PFME, özellikle stres tip idrar kaçırma ve hafif pelvik organ sarkması durumlarında en etkili ve en çok önerilen müdahale olarak öne çıkmaktadır. Biyogeribildirim, motor kontrolü ve egzersiz uyumu artırmakta; özellikle yaşlı bireylerde ve doğum sonrası kadınlarda faydalı olmaktadır. Elektrik stimülasyonu, istemli kas kasılımı zayıf olan hastalarda olumlu sonuçlar vermektedir. Manuel terapi, hipertonic kas aktivitesine bağlı pelvik ağrı olgularında etkili bulunmuştur. Mesane eğitimi ve tuvalet alışkanlıkları eğitimi gibi davranışsal stratejiler, tedavi sonuçlarını daha da iyileştirmektedir. Mobil uygulamalarla yürütülen PFME ve dijital araçlara yönelik kanıtlar giderek artmaktadır; ancak kullanıcı katılımı değişkenlik

Address for Correspondence: Asst. Prof. Özge Baykan Çopuroğlu, Kayseri University İncesu Ayşe and Saffet Arslan Health Services Vocational School, Department of Treatment and Rehabilitation, Physiotherapy Program, Kayseri, Türkiye

E-mail: ozgebaykancopuroglu@gmail.com

ORCID IDs of the authors: Ö.B.Ç.: 0000-0003-1014-8845, B.E.: 0000-0001-7212-2937

Cite this article as: Baykan Çopuroğlu Ö, Ergün B. Physiotherapy for pelvic dysfunction: a comprehensive review. Bezmialem Science. 2026;14(4):415-424

Received: 31.07.2025

Accepted: 10.11.2025

Published date: 26.10.2026



©Copyright 2026 Author(s). Published by Galenos Publishing House on behalf of Bezmialem Vakıf University. Licensed by Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0)

ABSTRACT

access to trained professionals, and integration of digital technologies are essential to improve treatment reach and long-term outcomes.

Keywords: Pelvic floor dysfunction, physiotherapy, pelvic floor muscle training, urinary incontinence, chronic pelvic pain

ÖZ

göstermektedir. Fizyoterapi, doğum sonrası, menopoz sonrası ve prostatektomi sonrası bireyler gibi özel popülasyonlarda da olumlu sonuçlar sağlamaktadır. Sonuç olarak, fizyoterapi PTD'nin yönetiminde kanıta dayalı ve bireyselleştirilmiş bir yaklaşım sunmaktadır. Tedaviye erişimin artırılması, uzman fizyoterapistlerin yaygınlaştırılması ve dijital teknolojilerin entegrasyonu, uzun vadeli sonuçların iyileştirilmesi açısından büyük önem taşımaktadır.

Anahtar Kelimeler: Pelvik taban disfonksiyonu, fizyoterapi, pelvik taban kas eğitimi, idrar kaçırma, kronik pelvik ağrı

Introduction

Pelvic floor dysfunction (PFD) is characterized by symptoms such as urinary and fecal incontinence, pelvic organ prolapse (POP), sexual dysfunction, and pelvic pain, which result from impairments in the function of pelvic muscles and connective tissues. Globally, PFD affects an estimated 25-45% of adult women, with prevalence varying by geography and socioeconomic status. Studies from low- and middle-income countries in Asia and Africa have reported comparable or even higher rates of urinary incontinence and pelvic pain compared to Western populations, underscoring the global burden of the disorder and the need for region-specific preventive strategies (1,2). A study conducted in Spain between 2021 and 2022 reported that more than half of the participating women experienced urinary incontinence, while 10.4% reported fecal incontinence, 14% had POP, and 18.7% reported pelvic pain. The study identified several risk factors associated with the development of PFD, including advanced age, gastrointestinal diseases, vaginal delivery, and instrumental birth (3,4). While historically underdiagnosed and undertreated, growing awareness of pelvic health has prompted significant interest in conservative management approaches, particularly physiotherapy.

The pelvic floor comprises a complex network of skeletal muscles, fascia, and ligaments, functioning collectively to support the pelvic organs, maintain continence, and contribute to lumbopelvic stability. Dysfunction may arise due to childbirth trauma, vaginal childbirth, aging, hormonal changes, surgery, obesity, high-impact sports, or chronic straining from constipation (5). However, PFD is not exclusive to women. Men may also experience PFD due to factors such as chronic prostatitis/chronic pelvic pain syndrome (CP/CPPS), sedentary lifestyle, obesity, excessive straining, and high-stress occupations in addition to post-prostatectomy incontinence (6).

Epidemiological studies estimate that nearly one in four adult women in the United States experiences at least one PFD, with prevalence increasing sharply after menopause (3). In contrast, data from regions such as sub-Saharan Africa and Southeast Asia remain limited, yet emerging

community-based surveys indicate that cultural stigma, limited access to physiotherapy, and lack of health education contribute to underdiagnosis and late presentation. These findings show the importance of global awareness and resource-adapted physiotherapy strategies.

Physiotherapy has emerged as the cornerstone of non-surgical intervention for PFD. Pelvic floor muscle training (PFMT), based on repetitive voluntary contractions of the PFMs, is recognized as a first-line treatment for various presentations, including stress urinary incontinence (SUI) and POP (7). Adjunct interventions such as biofeedback, electrical stimulation (ES), and bladder training enhance the effectiveness of PFMT. Through these approaches, healthcare professionals including physicians, physiotherapists, and midwives contribute to improvements in quality of life and reductions in symptoms (8,9). The mechanisms through which physiotherapy exerts benefit include neuromuscular re-education, increased muscle tone and endurance, and improved sensorimotor integration.

In recent years, robust evidence from randomized controlled trials and systematic reviews has confirmed the efficacy of physiotherapy in treating PFD. A Cochrane review encompassing over 30 trials demonstrated that women undergoing supervised PFMT were eight times more likely to report symptom improvement compared to controls (10). Another meta-analysis emphasized the cost-effectiveness of early physiotherapy intervention in preventing surgical management for SUI and POP (11). Moreover, newer approaches incorporating biofeedback and digital health platforms have shown promise in improving patient adherence and engagement, particularly in underserved populations (12). International clinical guidelines, including those from the International Continence Society, the European Association of Urology, and the American College of Obstetricians and Gynecologists, consistently recommend PFMT as the first-line conservative management for PFD.

Despite these advances, challenges remain. Many clinicians lack standardized training in pelvic health rehabilitation, and access to specialized physiotherapy services varies

by region. Therefore, a personalized, multidisciplinary approach grounded in patient education, structured assessment, and evidence-based protocols is essential to optimize outcomes in this complex population. This review aims to synthesize recent evidence (2014-2024) on physiotherapy interventions for pelvic dysfunction. We discuss the pathophysiology and classification of PFD, present a detailed overview of physiotherapeutic techniques, and evaluate clinical outcomes across various subgroups, including postpartum women, menopausal individuals, and male patients. By consolidating current knowledge, we hope to guide clinical practice and inform future research directions in pelvic floor rehabilitation.

The pelvic floor is a multilayered structure composed of skeletal muscles, ligaments, fascia, and neurovascular elements that support the pelvic organs and contribute to urinary and fecal continence, sexual function, and lumbopelvic stability (5). The primary muscular layer, the pelvic diaphragm, consists of the levator ani group—pubococcygeus, iliococcygeus, and puborectalis—and the coccygeus muscle. These muscles work synergistically with the abdominal wall and diaphragm to regulate intra-abdominal pressure and maintain core stability (13). Integrity of the pelvic floor can be compromised by vaginal delivery, hormonal changes, aging, obesity, or chronic increases in intra-abdominal pressure, leading to dysfunctions such as incontinence or prolapse (14). Decreased estrogen after menopause contributes to connective tissue weakening and muscle atrophy, which explains the rising prevalence of PFD with age. In summary, the pelvic floor is a biomechanically and neurologically complex system essential for urogenital and gastrointestinal health. Any disruption to its muscular, fascial, or neural components can result in dysfunction, emphasizing the need for physiotherapeutic interventions grounded in sound anatomical understanding.

PFD represents a heterogeneous group of disorders characterized by the inability of the pelvic floor muscles and connective tissue structures to perform their supportive and sphincteric roles adequately. The classification of PFD has evolved alongside advancements in imaging, diagnostics, and clinical assessment tools, enabling more precise identification of distinct phenotypes. Broadly, PFD can be categorized into three primary functional disturbances: underactive pelvic floor, overactive pelvic floor, and pelvic floor support defects (15).

Underactive PFD, the most commonly recognized subtype, is characterized by insufficient contraction strength or endurance of the pelvic floor muscles. It often manifests clinically as SUI, fecal incontinence, or POP. The pathogenesis of underactive PFD is frequently linked to mechanical trauma to the levator ani muscles or pudendal nerve, most commonly during vaginal childbirth (14).

Conversely, overactive PFD is characterized by increased resting tone or failure of the pelvic floor muscles to relax

appropriately. This condition is frequently associated with CPPSs, including dyspareunia, vulvodynia, and interstitial cystitis/bladder pain syndrome. Patients may present with urinary urgency, hesitancy, or incomplete voiding, often mistaken for urological pathology. Etiological factors include psychological stress, a history of sexual trauma, and central sensitization processes, wherein heightened nociceptive signaling alters pain perception and muscular tone (16). Electromyographic studies and palpation-based assessments frequently reveal myofascial trigger points and hypertonicity in affected individuals (17).

The third category, pelvic floor support defects, encompasses structural impairments such as cystocele, rectocele, enterocele, and uterine or vault prolapse. These are most commonly observed in women who have experienced vaginal childbirth but also occur following hysterectomy or pelvic surgery. Disruption or attenuation of the endopelvic fascia, uterosacral ligaments, or arcus tendinous fascia pelvis plays a central role in the pathophysiology of these conditions (18). Risk increases significantly with age, obesity, chronic coughing, connective tissue disorders such as Ehlers-Danlos syndrome, and occupational or lifestyle factors involving heavy lifting (19).

In both women and men, PFD may also be secondary to iatrogenic causes. Radical pelvic surgeries, including prostatectomy, hysterectomy, or bowel resection, frequently disrupt pelvic nerve pathways and impair muscular coordination. In men, the prevalence of post-prostatectomy urinary incontinence remains a significant concern, often linked to sphincter insufficiency or denervation of pelvic floor musculature (20).

Another increasingly recognized contributor to PFD is behavioral and postural dysfunction. Improper toileting behavior (e.g., habitual straining or delayed voiding), prolonged sitting, and poor core stabilization contribute to altered pelvic biomechanics, muscular imbalance, and neuromuscular discoordination (21). Moreover, psychosocial stress and anxiety disorders have been linked to increased pelvic floor muscle tone and dysregulation of the autonomic nervous system, further complicating the clinical picture (22).

In summary, PFD arises from an interplay of biomechanical, neurological, hormonal, and psychosocial factors. Its heterogeneity necessitates a thorough understanding of the underlying etiology to inform diagnosis and direct appropriate conservative treatment. Physiotherapy, as a frontline intervention, must be adapted to address the specific dysfunction subtype—whether hypotonus, hypertonus, or structural insufficiency—thereby ensuring optimal patient-centered care.

Clinical and Research Consequences

Although this review is narrative in nature, a structured literature search was conducted to ensure methodological

transparency. PubMed, Scopus, Web of Science, and Google Scholar were searched for articles published between January 2014 and December 2024 using the following keywords: “pelvic floor dysfunction”, “pelvic physiotherapy”, “pelvic floor muscle training”, “urinary incontinence”, “pelvic pain”. Studies were included if they: (1) involved adult populations; (2) evaluated physiotherapy interventions; and (3) were clinical trials, systematic reviews, meta-analyses or clinical guidelines. Non-English papers, case reports, pediatric-only studies, and articles without full text were excluded. Reference lists of relevant studies and guideline documents were additionally screened to identify further sources.

PFD encompasses a wide range of conditions that affect the function of the pelvic floor musculature and connective tissues, leading to symptoms such as urinary incontinence, POP, sexual dysfunction, and chronic pelvic pain. Physiotherapy plays a central role in the conservative management of these conditions by offering a set of tailored interventions that target muscular performance, neuromuscular control, tissue mobility, and behavioral habits. The diversity in pathophysiology across PFD subtypes necessitates a multifaceted treatment approach that often combines several physiotherapy modalities (15).

PFMT remains the gold standard for conservative management of hypotonic PFDs, particularly SUI and POP. PFMT consists of the voluntary contraction and relaxation of pelvic floor muscles with the aim of improving their strength, endurance, and coordination. A high-quality Cochrane review by Dumoulin et al. (7) showed that women who received PFMT were significantly more likely to report improvement in symptoms and quality of life compared to no treatment. In another meta-analysis, Wang et al. (11) demonstrated that PFMT also had a preventive effect on the progression of POP in women over a 12-month period. However, the majority of trials have evaluated outcomes only up to 6-12 months, and there remains a critical lack of evidence on the durability of PFMT benefits beyond this timeframe. Long-term follow-up studies (≥ 3 years) are urgently needed to determine whether improvements in continence, pelvic support, and pain control are sustained.

Another key limitation involves the wide heterogeneity in PFMT protocols across studies. Differences in training frequency (ranging from 1-3 sessions per week), intensity (maximal versus submaximal contractions), supervision levels (home-based versus supervised), and duration (4-24 weeks) make it difficult to establish universal clinical guidelines. This methodological variability highlights the need for standardized, evidence-based protocols to improve reproducibility and guide international recommendations (Table 1).

In cases of severe muscular weakness or partial denervation, ES can be utilized to initiate passive muscle contractions. ES delivers low-frequency impulses via vaginal or anal probes,

which activate motor units and improve recruitment of pelvic floor fibers. This modality is particularly useful in postpartum rehabilitation, post-surgical recovery, and for patients with neurogenic bladder. Lunardi et al. (23) demonstrated that ES combined with PFMT significantly improved continence outcomes in women with SUI compared to PFMT alone. Sarmiento et al. (24), emphasized its utility as a transitional tool for patients who were unable to generate sufficient voluntary contractions. A systematic review and meta-analysis by Chen et al. (25) also found benefits in early ES use following radical prostatectomy, contributing to faster return to continence.

For individuals presenting with overactive or hypertonic pelvic floor muscles—often manifesting as pelvic pain, vaginismus, or dyspareunia—manual therapy becomes a cornerstone of treatment. This includes techniques such as myofascial release, trigger point therapy, connective tissue mobilization, and neural gliding, applied either internally (via vaginal or rectal access) or externally. These methods aim to reduce muscle tone, desensitize painful tissue, and improve blood flow and mobility. Frederice et al. (17) reported that a series of manual therapy sessions significantly reduced pain severity and improved pelvic floor relaxation in women with vulvar pain syndromes. Murad-Regadas et al. (26) found similar results in patients with levator ani syndrome, with improvements in muscle tone and rectal pressure profiles. Moreover, Bittelbrunn et al. (27) showed the effectiveness of combining manual therapy with relaxation techniques in women suffering from chronic pelvic pain, showing over 70% reduction in symptom scores. Nevertheless, the evidence base for manual therapy is limited by small sample sizes, short intervention durations, and heterogeneous methodologies. Few randomized controlled trials compare manual therapy directly with other physiotherapy modalities, underscoring the need for large-scale, standardized trials to confirm its efficacy and optimize treatment protocols.

In addition to active and passive muscle-based approaches, behavioral therapies and bladder retraining protocols play a crucial role, particularly for patients with urge urinary incontinence, frequency, or dysfunctional voiding. These interventions involve educating the patient about optimal voiding habits, fluid intake, and urgency suppression strategies. Huang et al. (28) demonstrated that structured bladder training combined with PFMT resulted in significantly better symptom control than either intervention alone in older women. Similarly, Booth et al. (29), emphasized the benefit of bladder diaries and scheduled voiding in improving adherence and empowering patients to manage their symptoms independently. Bladder retraining is also essential in reducing detrusor overactivity and promoting central nervous system habituation.

Technological advancements have allowed pelvic floor physiotherapy to evolve beyond the clinic. Telerehabilitation and mobile app-based PFMT programs offer guided

exercise plans, symptom tracking, and remote therapist consultations. These platforms increase accessibility, particularly for patients in rural areas or those with mobility limitations. de Araujo et al. (12) conducted a randomized controlled trial comparing supervised physiotherapy to app-based PFMT and found no significant difference in outcomes after 12 weeks of training, though adherence was higher in the supervised group. Woodley et al. (30) reviewed digital health tools and concluded that mobile applications could be effective, especially when combined with periodic therapist feedback. Furthermore, Harper et al. (31) suggested that gamification and reminder systems embedded in apps enhanced user engagement and long-term compliance. However, real-world implementation of digital physiotherapy is often limited by barriers such as adherence variability, data privacy concerns, limited digital literacy among older adults, and inconsistent internet access in low-resource settings. Hybrid care models that combine in-person supervision with digital support may represent the most effective strategy for optimizing engagement and clinical outcomes.

A particularly promising direction in recent years has been the use of multimodal physiotherapy, where several interventions are combined based on patient presentation. For example, a 2018 prospective cohort study by Allaire et al. (32) demonstrated that a tailored program incorporating PFMT, manual therapy, cognitive-behavioral strategies, and mindfulness reduced chronic pelvic pain and improved sexual function in women over 12 weeks. Multimodal approaches reflect the growing recognition that PFD is a biopsychosocial condition, requiring interventions that address both physical and psychological dimensions of care (33,34).

Taken together, the physiotherapy management of PFD is grounded in a diverse arsenal of methods, each supported by clinical evidence and tailored to the functional status and goals of the patient. A successful rehabilitation program not only restores physical function but also improves self-efficacy, mental well-being, and quality of life. Table 2 summarizes the major evidence-based physiotherapy interventions and their clinical outcomes in PFD.

PFD affects diverse populations across the lifespan, and the approach to physiotherapy must be tailored to the specific physiological, anatomical, and psychosocial needs of each subgroup. Among the most clinically relevant populations are pregnant and postpartum women, peri- and postmenopausal women, and men, particularly those undergoing urological surgeries. Each group presents unique challenges and requires targeted assessment and intervention strategies supported by evolving clinical evidence. Emerging research also highlights the importance of early screening and preventive physiotherapy programs to minimize symptom progression in high-risk individuals.

Pregnant and postpartum women represent a high-risk population for the development of pelvic floor disorders due to mechanical, hormonal, and neuromuscular changes during pregnancy and childbirth. Increased intra-abdominal pressure, weight gain, and the effects of relaxing on connective tissue contribute to pelvic floor weakening, often leading to SUI and POP. Vaginal delivery, especially with prolonged labor, instrumental assistance, or perineal trauma, significantly increases the risk of levator ani avulsion and pudendal nerve injury. Several randomized controlled trials have demonstrated the preventive and therapeutic effects of antenatal and postnatal PFMT. A more recent meta-analysis by Woodley et al. (35) confirmed that postnatal PFMT significantly improved continence and pelvic muscle function, particularly when started within the first six weeks after delivery. Yet, data beyond one year postpartum remain scarce, highlighting the need for long-term follow-up to assess maintenance of muscle strength and continence outcomes.

Menopausal and postmenopausal women represent another population commonly affected by PFD, largely due to estrogen deficiency and connective tissue remodeling that accompany the hormonal changes of menopause. Decreased estrogen levels result in atrophy of the pelvic floor musculature and urogenital mucosa, leading to symptoms such as urge incontinence, vaginal dryness, POP, and sexual dysfunction. Physiotherapy in this population emphasizes not only PFMT but also strategies aimed at improving tissue elasticity, postural alignment, and pelvic

Table 1. Summary of pelvic floor muscle training (PFMT) protocol variations across studies

Author (year)	Frequency	Contraction parameters	Supervision type	Program duration
Dumoulin et al. (7) (2018)	3×/day	8-10 sec maximal contraction+relaxation	Supervised+home program	12 weeks
Woodley et al. (35) (2020)	Daily	5-8 sec contraction, 6-10 reps	Home-based	8-12 weeks
Wang et al. (11) (2022)	3 sessions/week	Progressive intensity	Supervised physiotherapy	16 weeks
Cacciari et al. (36) (2019)	2-3×/day	Slow (6-8 sec)+fast (1-2 sec) contractions	Mixed (clinic+home)	12 weeks
de Araujo et al. (12) (2024)	App-based daily program	Moderate-intensity contractions	Remote/digital	8-12 weeks
Starzec-Proserpio et al. (37) (2024)	2×/day	PFMT+relaxation/diaphragmatic breathing	Supervised multimodal	8 weeks
Chen et al. (25) (2024)	Daily	PFMT+electrical stimulation (for weak muscles)	Clinic-based	12 weeks

Table 2. Overview of evidence-based physiotherapy approaches in pelvic floor disorders

Intervention	Target PFD type	Clinical benefits	Key evidence
Pelvic floor muscle training (PFMT)	Stress urinary incontinence (SUI), pelvic organ prolapse (POP), postpartum dysfunction	Improves muscle strength, continence, and pelvic support; prevents POP progression	Dumoulin et al. (7), Wang et al. (11), Woodley et al. (35), Cacciari et al. (36)
Biofeedback training	SUI, postpartum, elderly	Enhances muscle awareness, control, and adherence to exercise programs	Nunes et al. (8), Leonardo et al. (9), Herderschee et al. (34), Ibrahim et al. (33)
Electrical stimulation	Neurogenic bladder, post-prostatectomy incontinence, severe weakness	Facilitates contraction in denervated muscles; accelerates continence recovery	Lunardi et al. (23), Sarmento et al. (24), Chen et al. (25), Johnson et al. (20)
Manual therapy	Chronic pelvic pain, hypertonic pelvic floor, vaginismus	Reduces trigger points, muscle tone, and pain; improves mobility and relaxation	Frederice et al. (17), Murad-Regadas et al. (26), Bittelbrunn et al. (27), Starzec-Proserpio et al. (37)
Behavioral therapy/bladder retraining	Urge UI, functional UI, frequency syndromes	Improves bladder control, reduces urgency and leakage; promotes adherence	Booth et al. (29), Huang et al. (28)
App-based and digital PFMT/telerehabilitation	Postpartum UI, elderly, rural or underserved populations	Expands access, supports adherence, and enables remote supervision	de Araujo et al. (12), Woodley et al. (30), Harper et al. (31)
Multimodal physiotherapy	Mixed-type dysfunctions, chronic pelvic pain	Integrates PFMT, manual therapy, mindfulness, and cognitive-behavioral methods for holistic outcomes	Allaire et al. (32), Bittelbrunn et al. (27), Starzec-Proserpio et al. (37)

PFD: Pelvic floor dysfunction

support mechanisms. Cacciari et al. (36) demonstrated that PFMT remains effective in postmenopausal women, particularly when delivered in a structured, supervised setting. Combining PFMT with hormone replacement therapy or topical estrogen may further enhance outcomes, though this warrants rigorous controlled trials.

Men with PFD constitute an increasingly recognized group, particularly in the context of post-prostatectomy incontinence and CPPS. Radical prostatectomy often results in partial denervation and weakening of the external urethral sphincter and pelvic floor, leading to significant urinary incontinence and reduced quality of life. Early initiation of PFMT—often combined with biofeedback or ES—has been shown to improve recovery of continence. Chen et al. (25) confirmed these findings in a randomized controlled trial, noting that men receiving physiotherapy interventions had significantly higher continence rates at 12 weeks post-op. However, evidence for the long-term (>1 year) effectiveness of physiotherapy in post-prostatectomy incontinence remains limited, and further studies are warranted. In addition to surgical recovery, men with CP/CPPS often present with hypertonic pelvic floor muscles. In such cases, physiotherapy focuses on manual therapy, down-training techniques, neuromuscular re-education, and lifestyle modifications. Starzec-Proserpio et al. (37) reported that 8 weeks of multimodal physiotherapy—

including manual therapy and breathing coordination—significantly reduced pain and urinary symptoms in men with CP/CPPS. These findings suggest that integrating physiotherapy into urological care pathways can improve holistic outcomes and patient satisfaction.

In addition to these primary populations, physiotherapists must also consider special cases such as athletes, transgender individuals, and those with connective tissue disorders such as Ehlers-Danlos syndrome. Athletes, especially those in high-impact sports, may experience PFD due to repetitive increases in intra-abdominal pressure. Culleton-Quinn et al. (38) highlighted that even asymptomatic athletes often show signs of pelvic floor overactivity, which can lead to pain, leakage, or prolapse later in life if unaddressed. Similarly, early evidence indicates that gender-affirming surgeries and hormone therapies may alter pelvic floor muscle function in transgender individuals, necessitating inclusive and specialized rehabilitation protocols.

Taken together, special populations affected by PFD demand a nuanced and evidence-informed physiotherapy approach. Factors such as age, hormonal status, type of delivery or surgery, comorbidities, and psychological readiness must be integrated into treatment planning. Ongoing research and clinical innovation continue to refine best practices in this field, with an increasing emphasis on

interdisciplinary collaboration and long-term management strategies. This evolving evidence base supports a shift from short-term symptom management toward sustained, preventive, and integrative models of pelvic health care.

The body of evidence reviewed in this article clearly affirms the role of physiotherapy as an effective, first-line intervention for a wide spectrum of PFD. PFMT, biofeedback, ES, manual therapy, and behavioral retraining are all supported by randomized controlled trials and systematic reviews that demonstrate improvements in continence, pain, quality of life, and functional outcomes across varied populations. However, the success of these interventions is heavily contingent on accurate diagnosis, individualized treatment plans, and patient adherence. Moreover, the sustainability of these improvements beyond one year remains uncertain, underscoring the need for longitudinal research to determine relapse rates, maintenance strategies, and cost-effectiveness over time.

While PFMT remains the cornerstone of conservative PFD management, the variability in training protocols, intensity, supervision, and outcome measures across studies presents challenges in establishing universal clinical guidelines. Dumoulin et al. (7) emphasized that supervised PFMT yielded superior results compared to unsupervised programs, yet access to trained pelvic health physiotherapists remains limited in many regions. Additionally, standardization of assessment tools, such as the PERFECT scheme or digital palpation scales, is inconsistent in clinical practice, which may hinder effective monitoring of patient progress and inter-study comparison. Future guideline development should focus on establishing minimum training standards, defining measurable outcome indicators, and encouraging multicenter collaborations to harmonize methodologies.

The integration of biofeedback and ES has expanded therapeutic possibilities, particularly for patients with poor proprioception or neuromuscular impairments. Nonetheless, some studies report marginal gains over PFMT alone, suggesting that these modalities may be most beneficial for specific subgroups rather than universally necessary (23,33). Cost-effectiveness, availability of equipment, and patient acceptability must also be considered when selecting adjunct interventions. Moreover, further high-quality studies are needed to elucidate long-term effects and optimal dosage parameters of electrotherapy in various PFD phenotypes. In this regard, future clinical trials should prioritize comparative effectiveness and pragmatic study designs that reflect real-world practice.

Manual therapy and down-training techniques have proven valuable for patients with pelvic floor overactivity and CPPSs. However, methodological heterogeneity in clinical trials, including varied treatment durations and therapist skill levels, limits generalizability. Frederice et al. (17) and Murad-Regadas et al. (26) provide strong evidence for the

role of internal myofascial release and trigger point therapy in symptom reduction, yet few large-scale randomized controlled trials have explored their comparative effectiveness or established standardized protocols. The biopsychosocial model, which incorporates psychological and behavioral dimensions into physiotherapy care, remains underutilized despite its growing support in chronic pain management literature. Incorporating this multidimensional framework into physiotherapy education and practice may improve both patient outcomes and clinician competency in complex pelvic pain syndromes.

Emerging technologies, particularly telehealth and mobile applications, offer promising avenues for expanding access to pelvic floor rehabilitation. However, while trials such as de Araujo et al. (12) confirm their clinical effectiveness, dropout rates and adherence challenges indicate that digital tools may not be suitable for all patients without additional therapist interaction. These findings highlight the need for hybrid models that combine in-person and digital care to meet diverse patient needs. Additionally, concerns remain regarding data privacy, accessibility among older populations, and the lack of personalized feedback in app-based programs. Regulatory frameworks should therefore ensure that telehealth platforms meet safety, confidentiality, and usability standards, while public health initiatives can promote digital literacy and equitable access.

Another critical aspect of this discussion lies in the management of special populations. Pregnant, postpartum, postmenopausal, and post-prostatectomy patients all require condition-specific modifications to physiotherapy protocols. While PFMT has proven effective across these subgroups, nuances such as hormonal status, surgical recovery timelines, and connective tissue integrity must guide intervention design (25,35). Furthermore, the needs of transgender individuals and those with rare connective tissue disorders remain significantly underrepresented in clinical trials. Addressing these gaps will require inclusive research designs, culturally competent care frameworks, and interdisciplinary collaboration between physiotherapists, gynecologists, urologists, and mental health professionals. Expanding participation of underrepresented groups in clinical trials will be essential to improve the external validity and global relevance of physiotherapy recommendations.

Despite the wealth of supportive data, barriers to implementation persist. Many healthcare systems lack reimbursement pathways for physiotherapy services related to pelvic health, contributing to underutilization and disparities in access. A study by Slade et al. (39) identified inadequate training during professional education and insufficient public awareness as additional obstacles to timely diagnosis and intervention. Addressing these systemic issues will be critical in optimizing population-level outcomes. From a policy perspective, national health

authorities should prioritize pelvic floor physiotherapy as a core component of preventive and rehabilitative healthcare. Structured postgraduate certification programs, continuing education opportunities, and inclusion of pelvic health modules in undergraduate physiotherapy curricula are necessary to standardize practice. Furthermore, establishing reimbursement mechanisms for conservative management could enhance accessibility and reduce the economic burden of surgical interventions.

Conclusion

In conclusion, physiotherapy offers a robust, evidence-based solution to PFD across diverse patient populations. The field continues to evolve with advances in clinical practice, technology, and interdisciplinary care. However, the translation of this evidence into sustainable health outcomes will depend on long-term data collection, policy integration, and equitable access. Future priorities should include (1) development of standardized PFMT and multimodal protocols; (2) creation of multidisciplinary pelvic health centers; (3) investment in telehealth infrastructure and digital literacy; and (4) implementation of national reimbursement frameworks. By aligning research, clinical training, and policy, physiotherapy can play a transformative role in improving global pelvic health and quality of life.

PFD is a prevalent condition that significantly impacts quality of life yet remains underdiagnosed and undertreated. Physiotherapy—including PFMT, biofeedback, ES, manual therapy, and behavioral strategies—offers a safe, effective, and evidence-based first-line approach across a range of PFD presentations.

When tailored to individual needs and delivered with clinical expertise, physiotherapy has demonstrated strong outcomes in improving continence, reducing pain, and enhancing functional quality of life in both women and men. Longitudinal evidence also suggests that early physiotherapy intervention can delay or even prevent surgical management for certain pelvic disorders, particularly when combined with lifestyle and behavioral modifications. Special populations, such as postpartum women, older adults, and prostatectomy patients, particularly benefit from targeted interventions.

Despite its proven efficacy, physiotherapy for pelvic health continues to face barriers related to limited-service access, inconsistent clinician training, and lack of standardized treatment protocols. Overcoming these challenges requires coordinated efforts among policymakers, educators, and healthcare providers. Establishing accredited pelvic health training programs, implementing reimbursement pathways for conservative interventions, and incorporating physiotherapy into multidisciplinary pelvic care models would enhance access and quality of care.

In parallel, emerging telehealth and app-based exercise systems hold promise for broadening reach, yet digital equity, privacy, and adherence must be carefully managed. Hybrid rehabilitation models—combining in-person assessment with digital follow-up—represent a pragmatic future direction for sustainable pelvic health programs.

Future research should focus on long-term effectiveness, real-world implementation, and inclusive strategies that address the needs of underserved and diverse populations. With sustained investment in education, infrastructure, and digital innovation, physiotherapy will continue to serve as a cornerstone of holistic, non-invasive pelvic health care and contribute substantially to global quality of life and healthcare sustainability.

Footnotes

Authorship Contributions

Concept: Ö.B.Ç., B.E., Design: Ö.B.Ç., B.E., Data Collection or Processing: Ö.B.Ç., Analysis or Interpretation: Ö.B.Ç., Literature Search: Ö.B.Ç., B.E., Writing: Ö.B.Ç., B.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. Bertuit J, Nzinga AL, Feipel V. Female urinary incontinence in Africa: prevalence estimates from a systematic review and meta-analysis. *Int Urogynecol J*. 2025;36:1901-20.
2. Nazzal Z, Khatib B, Al-Quqa B, Abu-Taha L, Jaradat A. The prevalence and risk factors of urinary incontinence among women with type 2 diabetes in the north West Bank: a cross-sectional study. *Lancet*. 2021;398(Suppl 1):S42.
3. Nygaard I, Barber MD, Burgio KL, Kenton K, Meikle S, Schaffer J, et al.; Pelvic Floor Disorders Network. Prevalence of symptomatic pelvic floor disorders in US women. *JAMA*. 2008;300:1311-6.
4. Peinado-Molina RA, Hernández-Martínez A, Martínez-Vázquez S, Rodríguez-Almagro J, Martínez-Galiano JM. Pelvic floor dysfunction: prevalence and associated factors. *BMC Public Health*. 2023;23:2005.
5. DeLancey JO. Pelvic floor anatomy and pathology. In: *Biomechanics of the female pelvic floor*. Academic Press, 2016. p.13-51.
6. Yani MS, Eckel SP, Kirages DJ, Rodriguez LV, Corcos DM, Kutch JJ. Impaired ability to relax pelvic floor muscles in men with chronic prostatitis/chronic pelvic pain syndrome. *Phys Ther*. 2022;102:pzac059.
7. Dumoulin C, Cacciari LP, Hay-Smith EJC. Pelvic floor muscle training versus no treatment, or inactive control treatments, for urinary incontinence in women. *Cochrane Database Syst Rev*. 2018;10:CD005654.
8. Nunes EFC, Sampaio LMM, Biasotto-Gonzalez DA, Nagano RCDR, Lucareli PRG, Politti F. Biofeedback for pelvic floor

- muscle training in women with stress urinary incontinence: a systematic review with meta-analysis. *Physiotherapy*. 2019;105:10-23.
9. Leonardo K, Seno DH, Mirza H, Afriansyah A. Biofeedback-assisted pelvic floor muscle training and pelvic electrical stimulation in women with overactive bladder: a systematic review and meta-analysis of randomized controlled trials. *Neurourol Urodyn*. 2022;41:1258-69. Erratum in: *Neurourol Urodyn*. 2023;42:889.
10. Dumoulin C, Hay-Smith J, Habée-Séguin GM, Mercier J. Pelvic floor muscle training versus no treatment, or inactive control treatments, for urinary incontinence in women: a short version Cochrane systematic review with meta-analysis. *Neurourol Urodyn*. 2015;34:300-8.
11. Wang T, Wen Z, Li M. The effect of pelvic floor muscle training for women with pelvic organ prolapse: a meta-analysis. *Int Urogynecol J*. 2022;33:1789-801.
12. de Araujo CC, Brito LGO, Marques AA, Juliato CRT. Use of a mobile application for pelvic floor muscle training in women with urinary incontinence: a randomized control trial. *Int Urogynecol J*. 2024;35:589-98.
13. Keizer A, Vandyken B, Vandyken C, Yardley D, Macedo L, Kuspinar A, et al. Predictors of pelvic floor muscle dysfunction among women with lumbopelvic pain. *Phys Ther*. 2019;99:1703-11.
14. Dietz HP. Pelvic floor trauma following vaginal delivery. *Curr Opin Obstet Gynecol*. 2006;18:528-37.
15. Bø K, Berghmans B, Mørkved S, Van Kampen M. Evidence-based physical therapy for the pelvic floor: Bridging science and clinical practice. 2nd ed. Amsterdam: Elsevier; 2021.
16. Thibault-Gagnon, S. Definitions and basic etiology of the overactive pelvic floor. In: *The overactive pelvic floor*. Cham: Springer International Publishing. 2016. p.1-16.
17. Frederice CP, Brito LGO, Pereira GMV, Lunardi ALB, Juliato CRT. Interventional treatment for myofascial pelvic floor pain in women: systematic review with meta-analysis. *Int Urogynecol J*. 2021;32:1087-96.
18. Deng ZM, Dai FF, Yuan MQ, Yang DY, Zheng YJ, Cheng YX. Advances in molecular mechanisms of pelvic organ prolapse (Review). *Exp Ther Med*. 2021;22:1009.
19. Shaw JM, Wolpern A, Wu J, Nygaard IE, Egger MJ. Postpartum sedentary behaviour and pelvic floor support: a prospective cohort study. *J Sports Sci*. 2023;41:141-50.
20. Johnson EE, Mamoulakis C, Stoniute A, Omar MI, Sinha S. Conservative interventions for managing urinary incontinence after prostate surgery. *Cochrane Database Syst Rev*. 2023;4:CD014799.
21. Hsu SL, Oda H, Shirahata S, Watanabe M, Sasaki M. Effects of core strength training on core stability. *J Phys Ther Sci*. 2018;30:1014-8.
22. Lamvu G, Carrillo J, Ouyang C, Rapkin A. Chronic pelvic pain in women: a review. *JAMA*. 2021;325:2381-91.
23. Lunardi AC, Foltran GC, Carro DF, Silveira LTY, Haddad JM, Ferreira EAG. Efficacy of electrical stimulation in comparison to active training of pelvic floor muscles on stress urinary incontinence symptoms in women: a systematic review with meta-analysis. *Disabil Rehabil*. 2025;47:3256-67.
24. Sarmiento ALC, Sá BS, Vasconcelos AG, Arcanjo DDR, Durazzo A, Lucarini M, et al. Perspectives on the therapeutic effects of pelvic floor electrical stimulation: a systematic review. *Int J Environ Res Public Health*. 2022;19:14035.
25. Chen R, Wang R, Yu Y, Zhao K, Li J. Intravaginal electrical stimulation for the treatment of pelvic floor dysfunction: a systematic review and meta-analysis. *Front Neurol*. 2024;15:1378494.
26. Murad-Regadas SM, Regadas FSP, de Oliveira MTCC, Vilarinho ADS, Fernandes G, Lima DMDR, et al. Is there a correlation between levator ani muscle defects, symptoms related to anorectal and urinary disorders, and anatomic and dynamic abnormalities in women with previous vaginal deliveries? *SN Compr Clin Med*. 2024;6:86.
27. Bittelbrunn CC, de Fraga R, Martins C, Romano R, Massaneiro T, Mello GVP, et al. Pelvic floor physical therapy and mindfulness: approaches for chronic pelvic pain in women-a systematic review and meta-analysis. *Arch Gynecol Obstet*. 2023;307:663-72.
28. Huang AJ, Chesney M, Schembri M, Raghunathan H, Vittinghoff E, Mendes WB, et al. Efficacy of a therapeutic pelvic yoga program versus a physical conditioning program on urinary incontinence in women : a randomized trial. *Ann Intern Med*. 2024;177:1339-49.
29. Booth J, Bliss D; International Continence Society Nursing Committee. Consensus statement on bladder training and bowel training. *Neurourology and Urodynamics*. 2020;39:1234-54.
30. Woodley SJ, Moller B, Clark AR, Bussey MD, Sangelaji B, Perry M, et al. Digital technologies for women's pelvic floor muscle training to manage urinary incontinence across their life course: scoping review. *JMIR Mhealth Uhealth*. 2023;11:e44929.
31. Harper RC, Sheppard S, Stewart C, Clark CJ. Exploring adherence to pelvic floor muscle training in women using mobile apps: scoping review. *JMIR Mhealth Uhealth*. 2023;11:e45947.
32. Allaire C, Williams C, Bodmer-Roy S, Zhu S, Arion K, Ambacher K, et al. Chronic pelvic pain in an interdisciplinary setting: 1-year prospective cohort. *Am J Obstet Gynecol*. 2018;218:114.e1-114.e12.
33. Ibrahim IK, Hameed MMA, Taher EM, Shaheen EM, Elsayy M. Efficacy of biofeedback-assisted pelvic floor muscle training in females with pelvic floor dysfunction. *Alexandria Journal of Medicine*. 2015;51:137-42.
34. Herderschee R, Hay-Smith EC, Herbison GP, Roovers JP, Heineman MJ. Feedback or biofeedback to augment pelvic floor muscle training for urinary incontinence in women: shortened version of a Cochrane systematic review. *Neurourol Urodyn*. 2013;32:325-9.
35. Woodley SJ, Lawrenson P, Boyle R, Cody JD, Mørkved S, Kernohan A, et al. Pelvic floor muscle training for preventing and treating urinary and faecal incontinence in antenatal and postnatal women. *Cochrane Database Syst Rev*. 2020;5:CD007471.
36. Cacciari LP, Dumoulin C, Hay-Smith EJ. Pelvic floor muscle training versus no treatment, or inactive control treatments, for urinary incontinence in women: a Cochrane systematic

- review abridged republication. *Braz J Phys Ther.* 2019;23:93-107.
37. Starzec-Proserpio M, Frawley H, Bo K, Morin M. (141) Multimodal physiotherapy for chronic pelvic pain in women—a systematic review and meta-analysis. *The Journal of Sexual Medicine.* 2024;21(Suppl 6): qdae161.116.
38. Culleton-Quinn E, Bø K, Fleming N, Mockler D, Cusack C, Daly D. Elite female athletes' experiences of symptoms of pelvic floor dysfunction: a systematic review. *Int Urogynecol J.* 2022;33:2681-711.
39. Slade SC, Hay-Smith J, Mastwyk S, Morris ME, Frawley H. Barriers and enablers to implementation of pelvic floor muscle training interventions for urinary incontinence. *Aust N Z Continence J.* 2018;24:118-25.

2026 Reviewer Index

Abdullah İLKTAÇ
Abdullah ŞİŞİK
Ahmet ÖZAYDIN
Akçahan GEPDİREMEN
Akile ZENGİN
Ali GUR
Ali Şefik KÖPRÜLÜ
Alper YENİGUN
Arda IŞIK
Ayşe ÇIRAKOĞLU
Ayşe KARAMAN
Ayşenur GÜNAYDIN AKYILDIZ
Banu KILIÇ
Bayram DOĞAN
Berza ŞEN YILMAZ
Betül Aycan UYSAL
Betül GÜVEN
Burak BUĞDAY
Burcu TOTUR DİKMEN
Burhan CEYLAN
Ceren DURMAZ ENGİN
Çağdaş YOKUŞOĞLU
Çağlar ÇETİN
Derya İPEKÇİOĞLU
Eda POLAT
Elif Dilara ARSLAN
Elifhan ALAGÖZ
Emine TEKİN
Eray Metin GÜLER
Erol ŞENTÜRK
Esmâ ŞENGENÇ
Esra EROĞLU ÖZKAN
Eyüp GEMİCİ
Fatih GEVHER
Gamze NALCI ÇALIK
Gülay OKAY
Gülezer AKPINAR

Hassen DAGHMOURA
Havva KOCAYİĞİT
Hayriye Esra ÜLKER
Hayrettin ÖZTÜRK
Hazal DUYAN YÜKSEL
Hüseyin KOCAASLAN
Hüsna ÖZVEREN
İlhan ONARAN
İşıl ÖZSOY SAYGIN
İsmihan İlknur UYSAL
Mahmut ULUGANYAN
Mehmet ÖZDİN
Mehmet Serkan KILIÇOĞLU
Mehmet TEKİN
Meliha AKSOY
Muhammed EKMEKYAPAR
Muhammed Halit TEKECİ
Muhammed Yunus BEKTAY
Muhammet HAMİTOĞLU
Mustafa AYATA
Mustafa ŞAHİN
Nilüfer KARAÇAY
Nimet SERMENLİ
Nuray KAHRAMAN AY
Nurgül YURTSEVEN
Ömer Faruk ÖZER
Samed ŞATIR
Savaş ÜSTÜNOVA
Seda Sezen GÖKTAŞ
Sevilay BATIBAY
Shamsi MEHDİYEV
Şahabettin SELEK
Tuba YILMAZER
Turgut DOLANBAY
Yasemin DURDU
Yazile SAYIN