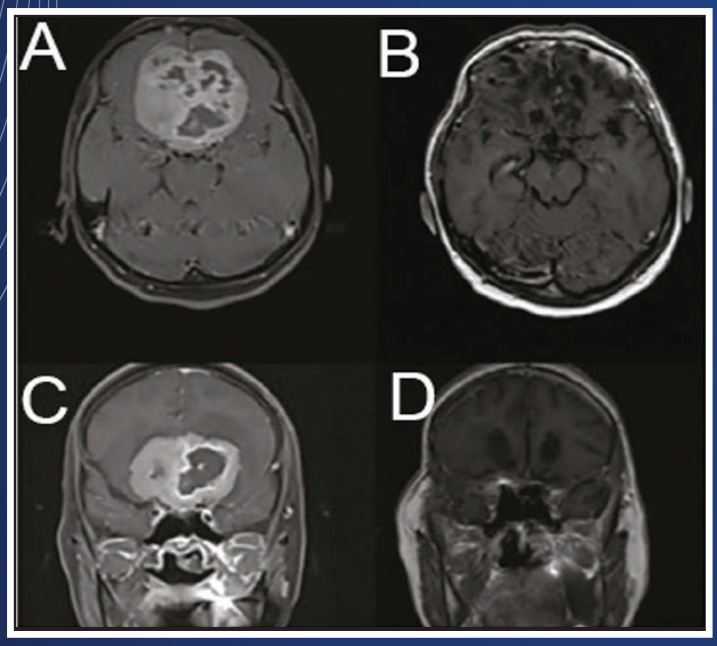
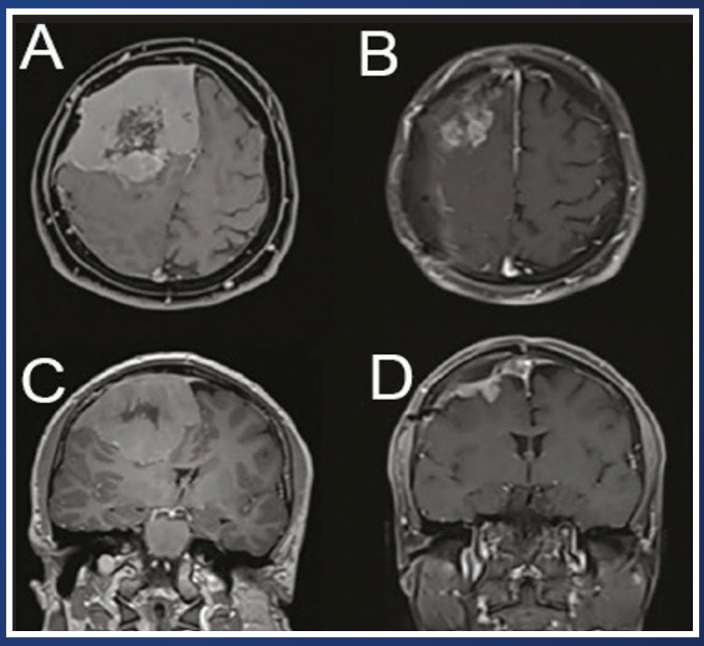
E-ISSN: 3023-6215

ANATOLIAN

JOURNAL of GENERAL MEDICAL RESEARCH

Volume/Cilt: 36 Issue/Sayı: 2 August/Ağustos 2026





galenos yayınevi

ANATOLIAN
JOURNAL of GENERAL MEDICAL RESEARCH

anatolianjmed.org

EDITORIAL BOARD

Editor in Chief

Savaş YAKAN, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of General Surgery, İzmir, Türkiye
savasyakan@gmail.com

Editors

İbrahim ÇUKUROVA, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Ear, Nose and Throat, İzmir, Türkiye
cukurova57@gmail.com

Pınar GENÇPINAR, Prof. MD, PhD
İzmir Katip Çelebi University Faculty of Medicine, Department of Pediatric Neurology, İzmir, Türkiye
drpinargencpinar@yahoo.com

Managing Editors/Section Editors

Mehmet ÜSTÜN, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of General Surgery, İzmir, Türkiye
dr.m.ustun@gmail.com

Ferhat DEMİRCİ, Assoc. Prof. MD, PhD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Medical Biochemistry, İzmir, Türkiye
Dokuz Eylül University, Institute of Health Sciences, Department of Neurosciences, İzmir, Türkiye
drdemirci05@gmail.com

Ozan Barış NAMDAROĞLU, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of General Surgery, İzmir, Türkiye
ozannamdaroglu@yahoo.com

Özlem YAĞIZ AGAYAROV, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Otorhinolaryngology and Head and Neck Surgery, İzmir, Tepecik
drozlemyagizagayarov@gmail.com

Gamze SARIKAYA UZAN, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Child Health and Diseases, Division of Pediatric Neurology, İzmir, Türkiye
gamzeuzan36@gmail.com

Burcu KAYHAN TETİK, Prof. MD
İnönü University Faculty of Medicine, Department of Family Medicine, Malatya, Türkiye
burcu.tetik@inonu.edu.tr

Belgin COŞKUN, Assoc. Prof. MD
University of Health Sciences Türkiye, Ankara Bilkent City Hospital, Department of Infectious Diseases and Clinical Microbiology, Ankara, Türkiye
belgintekin@yahoo.com

Metin ÇOKSEVİM, Assoc. Prof. MD
Ondokuz Mayıs University Faculty of Medicine, Department of Cardiology, Samsun, Türkiye
metincoksevim@gmail.com

Statistic Editor

Ferhan ELMALI, Prof. MD
İzmir Katip Çelebi University Faculty of Medicine, Department of Biostatistics, İzmir, Türkiye
elmaliferhan@yahoo.com

Publishing Secretary

Hülya ERGEN
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Nursing, İzmir, Türkiye
hulyaergen1@hotmail.com

Yusuf ARIKAN
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Urology, İzmir, Türkiye
dryusufarikan@gmail.com

Gülberat TOTUR
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Pediatrics, İzmir, Türkiye
dr_ginceg@hotmail.com
gulberat.incetotur@sbu.edu.tr

Founder

Savaş YAKAN, Prof. MD

English Language Editor

Galenos Publishing House

The Anatolian Journal of General Medical Research is the official journal of the İzmir Tepecik Education and Research Hospital.
It is published three times a year (April, August, December).
Anatolian Journal of General Medical Research is an open access, free and peer-reviewed journal and indexed in
Scopus, idealonline, J-Gate, Sudoc, Gale, Türkiye Atıf Dizini, TR Dizin, Türk Medline and **EBSCO.**

©All rights are reserved. Rights to the use and reproduction, including in the electronic media, of all communications, papers, photographs and illustrations appearing in this journal belong to the Anatolian Journal of General Medical Research. Reproduction without prior written permission of part or all of any material is forbidden. The journal complies with the Professional Principles of the Press.

galenos yayınevi

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

Online Publication Date: August 2026

E-ISSN: 3023-6215

International periodical journal published three times in a year.

ANATOLIAN
JOURNAL of GENERAL MEDICAL RESEARCH

ADVISORY BOARD

anatolianjmed.org

Zehra Hilal ADIBELLİ, Assoc. Prof.
University of Health Sciences Türkiye, Bozyaka Health Application and Research Center, Department of Radiology, İzmir, Türkiye

Cezmi AKKIN, Prof. MD
Ege University Faculty of Medicine, Department of Ophthalmology, İzmir, Türkiye

Ekin Özgür AKTAŞ, Prof. MD
Ege University Faculty of Medicine, Department of Forensic Medicine, İzmir, Türkiye

Safiye AKTAŞ, Prof. MD
Dokuz Eylül University, Institute of Oncology, Department of Basic Oncology, İzmir, Türkiye

Ayşe Berna ANIL, Prof. MD
University of Health Sciences Türkiye, Tepecik Health Application and Research Center, Department of Pediatrics, İzmir, Türkiye

Murat ANIL, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Pediatrics, İzmir, Türkiye

Fazıl APAYDIN, Prof. MD
Ege University Faculty of Medicine, Department of Otorhinolaryngology, İzmir, Türkiye

İlker Burak ARSLAN, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Otorhinolaryngology, İzmir, Türkiye

Suna ASILSOY, Prof. MD
Dokuz Eylül University Faculty of Medicine, Department of Pediatrics, Division of Pediatric Allergy, İzmir, Türkiye

Arzu AVCI, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Medical Pathology, İzmir, Türkiye

Cengiz AYDIN, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of General Surgery, İzmir, Türkiye

Semin AYHAN, Prof. MD
Celal Bayar University Faculty of Medicine, Department of Medical Pathology, Manisa, Türkiye

Ali Rahmi BAKİLER, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Pediatric Cardiology, İzmir, Türkiye

Selen BAYRAKTAROĞLU, Assoc. Prof. MD
Ege University Faculty of Medicine, Department of Radiology, İzmir, Türkiye

Oktay BİLGİ, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Bozyaka Health Application and Research Center, Department of Hematology, İzmir, Türkiye

Giray BOZKAYA, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Bozyaka Health Application and Research Center, Department of Pediatric Genetics, İzmir, Türkiye

Pervin BOZKURT, Prof. MD
İstanbul University-Cerrahpaşa, Faculty of Medicine, Department of Anesthesiology and Reanimation, İstanbul, Türkiye

Demet CAN, Prof. MD
Balıkesir University Faculty of Medicine, Department of Pediatrics, Balıkesir, Türkiye

Gül CANER MERCAN, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Otorhinolaryngology, İzmir, Türkiye

Cengiz CEYLAN, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Hematology, İzmir, Türkiye

İzzetiye Ebru ÇAKIR, Prof. MD
İzmir Katip Çelebi University Faculty of Medicine, Department of Medical Pathology, İzmir, Türkiye

Şebnem ÇALKAVUR, Assoc. Prof. MD
University of Health Sciences Türkiye, Dr. Behçet Uz Education and Research Hospital, Department of Neonatology, İzmir, Türkiye

Taner ÇAMSARI, Assoc. Prof. MD
Dokuz Eylül University Faculty of Medicine, Department of Nephrology, İzmir, Türkiye

Tanju ÇELİK, Assoc. Prof. MD
University of Health Sciences Türkiye, Dr. Behçet Uz Education and Research Hospital, Department of Pediatrics, İzmir, Türkiye

Ayfer ÇOLAK, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Medical Biochemistry, İzmir, Türkiye

İbrahim ÇUKUROVA, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Otorhinolaryngology, İzmir, Türkiye

Füsun DEMİRÇİVİ ÖZER, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Neurosurgery, İzmir, Türkiye

Tevfik DORAK, Prof. MD
Kingston University Faculty of Science, Department of Engineering and Computing, London, UK

Nuket ELİYATKIN, Assoc. Prof. MD
Adnan Menderes University Faculty of Medicine, Department of Medical Pathology, Aydın, Türkiye

Hülya ELLİDOKUZ, Prof. MD
Dokuz Eylül University Faculty of Medicine, Department of Public Health, İzmir, Türkiye

Erhan ESER, Prof. MD
Celal Bayar University Faculty of Medicine, Department of Public Health, Manisa, Türkiye

Demet ETİT, Assoc. Prof. MD
İzmir Katip Çelebi University Faculty of Medicine, Department of Medical Pathology, İzmir, Türkiye

Ferah GENEL, Prof. MD
University of Health Sciences Türkiye, Dr. Behçet Uz Education and Research Hospital, Department of Pediatric Immunology and Allergy, İzmir, Türkiye

Ahmet GÜL, Prof. MD
İstanbul University, İstanbul Faculty of Medicine, Department of Rheumatology, İstanbul, Türkiye

İbrahim GÜLHAN, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Obstetrics and Gynecology, İzmir, Türkiye

Mehmet HELVACI, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Pediatrics, İzmir, Türkiye

ANATOLIAN
JOURNAL of GENERAL MEDICAL RESEARCH

ADVISORY BOARD

anatolianjmed.org

Münevver HOŞGÖR, Prof. MD
University of Health Sciences Türkiye, Dr. Behçet Uz Education and Research Hospital, Department of Pediatric Surgery, İzmir, Türkiye

Yusuf Özlem İLBEY, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Urology, İzmir, Türkiye

Tolga İNCE, Assoc. Prof. MD
Dokuz Eylül University Faculty of Medicine, Department of Pediatrics, Division of Social Pediatrics, İzmir, Türkiye

Sema KALKAN UÇAR, Prof. MD
Ege University Faculty of Medicine, Department of Pediatrics, İzmir, Türkiye

Ali KANIK, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Pediatrics, İzmir, Türkiye

Levent KARAPINAR, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Orthopedics and Traumatology, İzmir, Türkiye

Belde KASAP DEMİR, Prof. MD
İzmir Katip Çelebi University Faculty of Medicine, Department of Pediatrics, İzmir, Türkiye

Ahmet KAYA, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Orthopedics and Traumatology, İzmir, Türkiye

Dayimi KAYA, Prof. MD
Dokuz Eylül University Faculty of Medicine, Department of Cardiology, İzmir, Türkiye

Aykut KEFİ, Prof. MD
Dokuz Eylül University Faculty of Medicine, Department of Urology, İzmir, Türkiye

Mehmet Zeynel KESKİN, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Urology, İzmir, Türkiye

Yasemin KILIÇ ÖZTÜRK, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Family Medicine, İzmir, Türkiye

Barış KILIÇASLAN, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Cardiology, İzmir, Türkiye

Cenk KIRAKLI, Assoc. Prof. MD
University of Health Sciences Türkiye, Suat Seren Pulmonary Diseases and Thoracic Surgery Health Application and Research Center, Department of Pulmonary Diseases, İzmir, Türkiye

Tayfun KİRAZLI, Prof. MD
Ege University Faculty of Medicine, Department of Otorhinolaryngology, İzmir, Türkiye

Gökhan KÖYLÜOĞLU, Prof. MD
İzmir Katip Çelebi University Faculty of Medicine, Department of Pediatric Surgery, İzmir, Türkiye

Şükran KÖSE, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Infectious Diseases, İzmir, Türkiye

Semra HIZ, Prof. MD
Dokuz Eylül University Faculty of Medicine, Department of Pediatrics, Division of Pediatric Neurology, İzmir, Türkiye

Nilgün KÜLTÜRSAY, Prof. MD
Ege University Faculty of Medicine, Department of Pediatrics, Division of Neonatology, İzmir, Türkiye

Timur MEŞE, Assoc. Prof. MD
University of Health Sciences Türkiye, Dr. Behçet Uz Education and Research Hospital, Department of Pediatric Cardiology, İzmir, Türkiye

Fatma MUTLUBAŞ, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Pediatric Nephrology, İzmir, Türkiye

Nazmi NARİN, Prof. MD
İzmir Katip Çelebi University Faculty of Medicine, Department of Pediatrics, Division of Pediatric Cardiology, İzmir, Türkiye

Nur OLGUN, Prof. MD
Dokuz Eylül University, Institute of Oncology, Department of Clinical Oncology, İzmir, Türkiye

Mustafa OLGUNER, Prof. MD
Dokuz Eylül University Faculty of Medicine, Department of Pediatric Surgery, İzmir, Türkiye

Özgür OLUKMAN, Assoc. Prof. MD
Çiğli Training and Research Hospital, Clinic of Neonatology, İzmir, Türkiye

Resmiye ORAL, Prof. MD
University of Iowa Stead Family, Department of Pediatrics, USA

Yeşim OYMAK, Assoc. Prof. MD
University of Health Sciences Türkiye, Dr. Behçet Uz Education and Research Hospital, Department of Pediatric Hematology and Oncology, İzmir, Türkiye

Dilek ÖNCEL, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Radiology, İzmir, Türkiye

Mehmet Yekta ÖNCEL, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Neonatology, İzmir, Türkiye

Hale ÖREN, Prof. MD
Dokuz Eylül University Faculty of Medicine, Department of Pediatric Hematology and Oncology, İzmir, Türkiye

Öner ÖZDOĞAN, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Cardiology, İzmir, Türkiye

Erdener ÖZER, Prof. MD
Dokuz Eylül University Faculty of Medicine, Department of Medical Pathology, İzmir, Türkiye

Mehmet ÖZEREN, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Obstetrics and Gynecology, İzmir, Türkiye

Nuray ÖZGÜLNAR, Prof. MD
İstanbul University, İstanbul Faculty of Medicine, Department of Public Health, İstanbul, Türkiye

Deniz ÖZTEKİN, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Obstetrics and Gynecology, İzmir, Türkiye

Mehmet Burak ÖZTOP, MD
İzmir Provincial Health Directorate, General Surgery Specialist, İzmir, Türkiye

Emel Ebru PALA, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Medical Pathology, İzmir, Türkiye

ANATOLIAN
JOURNAL of GENERAL MEDICAL RESEARCH

ADVISORY BOARD

anatolianjmed.org

İbrahim PİRİM, Prof. MD
İzmir Katip Çelebi University Faculty of Medicine, Department of Medical Biology and Genetics, İzmir, Türkiye

Münire Ece SABAH, Prof. MD
Ege University Faculty of Dentistry, Department of Orthodontics, İzmir, Türkiye

Muzaffer SANCI, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Obstetrics and Gynecology, İzmir, Türkiye

Erkin SERDAROĞLU, Prof. MD
University of Health Sciences Türkiye, Dr. Behçet Uz Education and Research Hospital, Department of Pediatric Nephrology, İzmir, Türkiye

Sevil SAYHAN, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Medical Pathology, İzmir, Türkiye

Caroline SEWRY, Prof. MD
UCL Queen Square Institute of Neurology, United Kingdom

Süheyla SÜRÜCÜOĞLU, Prof. MD
Celal Bayar University Faculty of Medicine, Department of Medical Microbiology, Manisa, Türkiye

Aydın ŞENCAN, Prof. MD
Celal Bayar University Faculty of Medicine, Department of Pediatric Surgery, Manisa, Türkiye

Erkan ŞENGÜL, Assoc. Prof. MD
University of Health Sciences Türkiye, Derince Health Application and Research Center, Department of Nephrology, Kocaeli, Türkiye

Mehmet ŞENEŞ, Assoc. Prof. MD
University of Health Sciences Türkiye, Ankara Health Application and Research Center, Department of Medical Biotechnology, Ankara, Türkiye

Mehmet TANRISEV, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Nephrology, İzmir, Türkiye

Ercüment TARCAN, Assoc. Prof. MD
İzmir Atatürk Training and Research Hospital, Clinic of General Surgery, İzmir, Türkiye

Ali TAYLAN, Assoc. Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Rheumatology, İzmir, Türkiye

Hasan TEKGÜL, Prof. MD
Ege University Faculty of Medicine, Department of Pediatrics, Division of Pediatric Neurology, İzmir, Türkiye

Hasan TEZER, Prof. MD
Pediatric Infectious Diseases, Private, Ankara, Türkiye

Dilek TOPRAK, Prof. MD
Namık Kemal University Faculty of Medicine, Department of Family Medicine, Tekirdağ, Türkiye

Tuba TUNCEL, Prof. MD
İzmir Katip Çelebi University Faculty of Medicine, Department of Pediatrics, İzmir, Türkiye

Ayşen TÜREDİ, Assoc. Prof. MD
Celal Bayar University Faculty of Medicine, Department of Pediatrics, Manisa, Türkiye

Meltem UYAR, Prof. MD
Ege University Faculty of Medicine, Department of Algology, İzmir, Türkiye

Nurettin ÜNAL, Prof. MD
Dokuz Eylül University Faculty of Medicine, Department of Pediatrics, Division of Pediatric Cardiology, İzmir, Türkiye

Aycan ÜNALP, Prof. MD
University of Health Sciences Türkiye, Dr. Behçet Uz Education and Research Hospital, Department of Children's Neurology, İzmir, Türkiye

Canan VERGİN, Assoc. Prof. MD
University of Health Sciences Türkiye, Dr. Behçet Uz Education and Research Hospital, Department of Pediatric Hematology and Oncology, İzmir, Türkiye

Raşit Vural YAĞCI, Prof. MD
Ege University Faculty of Medicine, Department of Pediatrics, Division of Pediatric Gastroenterology and Hepatology, İzmir, Türkiye

Mehmet YALAZ, Prof. MD
Ege University Faculty of Medicine, Department of Pediatrics, İzmir, Türkiye

Önder YAVAŞCAN, Prof. MD
Medipol University Faculty of Medicine, Department of Pediatrics, Division of Pediatric Nephrology, İstanbul, Türkiye

Mehmet YILDIRIM, Assoc. Prof. MD
University of Health Sciences Türkiye, Bozyaka Health Application and Research Center, Department of General Surgery, İzmir, Türkiye

Dilek YILMAZ ÇİFTDOĞAN, Prof. MD
University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Pediatric Infectious Diseases, İzmir, Türkiye

İsmail YILMAZ, Prof. MD
İzmir Katip Çelebi University Faculty of Medicine, Department of Medical Pharmacology, İzmir, Türkiye

Orhan YILMAZ, Assoc. Prof. MD
University of Health Sciences Türkiye, Diskapi Yildirim Beyazit Health Application and Research Center, Department of Otorhinolaryngology, Ankara, Türkiye

Murat Muhtar YILMAZER, Assoc. Prof. MD
University of Health Sciences Türkiye, Dr. Behçet Uz Education and Research Hospital, Department of Pediatrics, İzmir, Türkiye

Seyran YİĞİT, Assoc. Prof. MD
İzmir Atatürk Training and Research Hospital, Clinic of Medical Pathology, İzmir, Türkiye

Please refer to the journal's webpage (https://tepecikdergisi.com/) for "Aims and Scope", "Instructions for Authors" and "Ethical Policy".

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

The Anatolian Journal of General Medical Research is indexed in **Scopus, idealonline, J-Gate, Sudoc, Gale, Türkiye Atıf Dizini, TR Dizin, Türk Medline, EBSCO.**

ANATOLIAN
JOURNAL of GENERAL MEDICAL RESEARCH

CONTENTS / İÇİNDEKİLER

anatolianjmed.org

LETTER TO THE EDITOR / EDİTÖRE MEKTUP

CLINICAL RESEARCH / KLİNİK ARAŞTIRMALAR

ANATOLIAN
JOURNAL of GENERAL MEDICAL RESEARCH

CONTENTS / İÇİNDEKİLER

anatolianjmed.org

DOI: 10.4274/anatoljmed.2026.63383
Anatol J Gen Med Res 2026;36(2):100-3

LETTER TO THE EDITOR / EDİTÖRE MEKTUP

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Pediatric Refractory Anti-GAD65 Positive Autoimmune Encephalitis Successfully Treated with Rituximab

Rituximab ile Başarıyla Tedavi Edilen Pediyatrik Dirençli Anti-GAD65 Pozitif Otoimmün Ensefalit

Gizem Doğan[1], Ökkeş Özgür Mart[2], Sunay Sibel Karayol[3], Yiğithan Güzin[1], Figen Baydan[1], Gamze Sarıkaya Uzan[1,4]

[1]University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Pediatric Neurology, İzmir, Türkiye
[2]University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Pediatric Intensive Care, İzmir, Türkiye
[3]University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Radiology, İzmir, Türkiye
[4]University of Health Sciences Türkiye, İzmir Faculty of Medicine, Department of Pediatric Neurology, İzmir, Türkiye

Cite as: Doğan G, Mart ÖÖ, Karayol SS, Güzin Y, Baydan F, Sarıkaya Uzan G. Pediatric refractory anti-GAD65 positive autoimmune encephalitis successfully treated with rituximab. Anatol J Gen Med Res. 2026;36(2):100-3

Keywords: Anti-GAD65, autoimmune encephalitis, rituximab, status epilepticus
Anahtar Kelimeler: Anti-GAD65, otoimmün ensefalit, rituksimab, status epileptikus

Dear Editor,

Autoimmune encephalitis (AE) is a group of encephalitides caused by an autoimmune response targeting antigens of the central nervous system. The incidence in the pediatric population is 1/100.000. AE has a wide spectrum of clinical manifestations. Children with AE may develop significant neurological symptoms within a short period, including cognitive dysfunction, behavioral changes, movement disorders, seizures, and disturbances of consciousness[1,2]. The diagnosis of AE is supported by clinical findings, neuroimaging, cerebrospinal fluid (CSF) analysis, and antibody tests. Recently, many AE-associated antibodies have been identified; anti-glutamic acid decarboxylase 65 (GAD65) is one of them. GAD65 antibodies are associated with neurological conditions such as stiff-person syndrome, AE, and cerebellar ataxia[3]. Here, we will present an anti-GAD65-positive AE case that developed into super-refractory status epilepticus (SE) while being followed for focal epilepsy and which we successfully treated with rituximab.

A previously healthy 18-month-old female patient presented with a focal tonic seizure. Prenatal and postnatal histories were unremarkable. Neurodevelopment was normal. At 18 months of age, brain magnetic resonance imaging (MRI) was normal, and electroencephalography (EEG) showed focal spike-and-slow-wave discharges in the right temporal area. Follow-up MRI at 9 years of age showed focal cortical dysplasia (FCD) involving the right frontoparietal region (Figure 1A). Valproic acid was given as the first anti-seizure medication (ASM). The patient continued to have focal seizures during follow-up; therefore, carbamazepine, lamotrigine, levetiracetam, and topiramate were added.

Address for Correspondence/Yazışma Adresi: Gizem Doğan, MD, University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Pediatric Neurology, İzmir, Türkiye
E-mail: gizemdoganmd@gmail.com
ORCID ID: orcid.org/0000-0002-2913-717X

Received/Geliş tarihi: 20.04.2026
Accepted/Kabul tarihi: 05.07.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

At 14 years of age, she presented to the emergency department with seizures. Seizure semiology was characterized by right-sided myoclonic-clonic movements of the face and arm, with impaired consciousness. Due to ongoing seizures, the patient was diagnosed with SE and managed according to current protocol[4]. Brain MRI performed during this period showed an edematous appearance in the right cerebral hemisphere and hyperintensity in the right occipital cortical area on T2A and fluid-attenuated inversion recovery (FLAIR) sequences (Figure 1B, C). The EEG performed on the first day showed epileptiform discharges originating from the temporal region of the right hemisphere, and these discharges occasionally spread intrahemispherically. A few days later, EEG monitoring in the pediatric intensive care unit revealed epilepsia partialis continua (Figure 2).

The patient had seizures that could not be controlled with ASM for more than 24 hours. Follow-up was initiated for the patient receiving mechanical ventilation. Therefore, it was classified as super-refractory SE. Because of suspected autoimmune involvement, intravenous immunoglobulin (IVIg) was administered at a dose of 1 g/kg. The patient responded well to this treatment, with seizures decreasing by more than 50%. A few days later, because of persistent seizures, pulse methylprednisolone therapy was initiated at 1.000 mg/day for 10 days. After this treatment, there was an improvement in consciousness and the general condition. We added lacosamide at 4 mg/kg/day to the treatment regimen to benefit from its effects on focal seizures. CSF AE panel was unremarkable, which includes AMPA1, AMPA2, anti-GABA B, anti-DPPX, anti-NMDA, anti-CASPR, and anti-LGI1 antibodies. The CSF AE panel, which included antibodies to AMPA1, AMPA2, GABA B, DPPX, NMDA, CASPR, and LGI1, was unremarkable. Thyroid and celiac

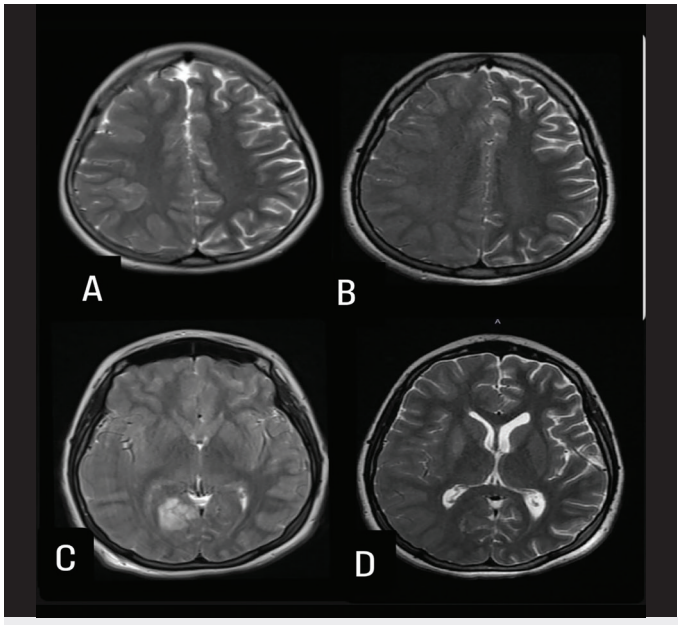


Figure 1. A) Cortical thickening and dysgyria are observed in the right frontoparietal area. This is considered to be focal cortical dysplasia. B, C) The right hemisphere of the brain appears edematous compared with the left hemisphere. Hyperintensity is observed in the right occipital cortex on T2A and FLAIR sequences. D) On the control MRI performed after immunotherapy, we observed that the hyperintensity on T2 and FLAIR imaging in the right occipital region disappeared

FLAIR: Fluid-attenuated inversion recovery, MRI: Magnetic resonance imaging

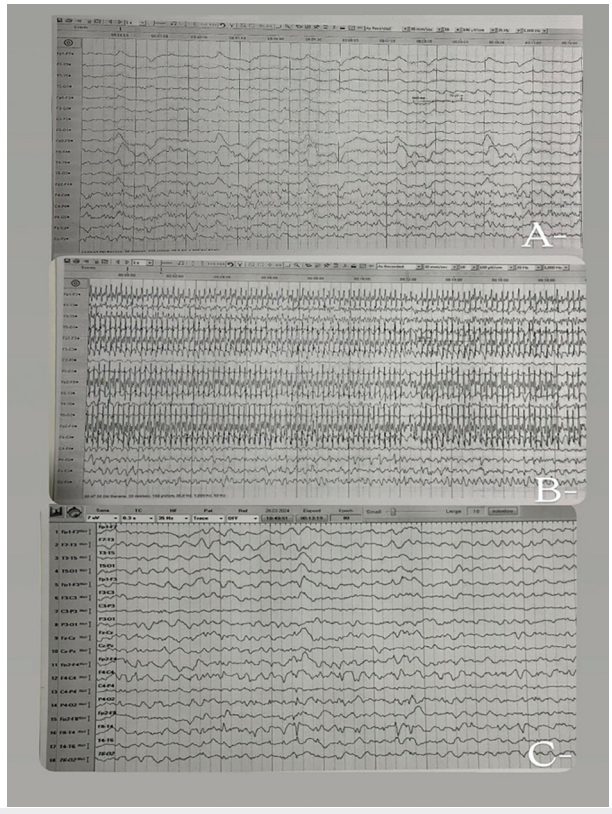


Figure 2. EEG images. A) Asymmetry between the right and left hemispheres is observed. The frequency and amplitude are higher in the right hemisphere than in the left hemisphere. B) Ictal EEG recording was performed during status epilepticus while intubated. C) After the rituximab treatment, phase-matching spike slow wave discharges are observed in the area corresponding to the right T4 electrode

EEG: Electroencephalography

antibodies, myelin oligodendrocyte glycoprotein antibody, aquaporin-4 antibody, ANA, and anti-dsDNA were negative. Serum anti-GAD65 antibody, measured by enzyme-linked immunosorbent assay, was 35 IU/mL (normal range <10 IU/mL). Other CSF studies, including glucose, protein, leukocyte count, culture, polymerase chain reaction, cytology, and oligoclonal bands, were unremarkable. CSF GAD65 antibody testing was planned for definitive diagnosis; however, it could not be performed due to financial constraints and limitations in health insurance coverage.

The patient was diagnosed with probable anti-GAD65-positive AE. We decided to administer rituximab to the patient who experienced recurrent seizures after a temporary seizure-free period following steroid treatment. Comprehensive evaluations were performed, including serological testing, measurement of Ig levels, lymphocyte panel, vaccination status, and chest X-ray. She had no active, serious infection that would prevent her from receiving rituximab. Then, we administered rituximab 375 mg/m^2 weekly for 4 weeks. After the first dose, we achieved complete resolution of seizures, and she was extubated. She was free from seizures after four weeks of treatment. We planned monthly IVIg treatment for the patient for at least 6 months, with rituximab treatment to commence after that period. On the control MRI, we observed that the hyperintensity in the right occipital area on T2 and FLAIR sequences had resolved. (Figure 1D). She was discharged in good general condition with spontaneous breathing and enrolled in a home-based rehabilitation and physical therapy program.

The diagnostic process for AE in children is challenging due to the wide clinical spectrum. Antibody detection plays a critical role in this process. Anti-GAD65 AE should be considered in the evaluation of all suspected cases. Seizures are typically refractory in anti-GAD65-positive AE. Steroids, IVIg, and plasmapheresis are first-choice treatments. If there is no response to first-line treatments, different treatment options should be considered. Rituximab is an anti-CD20 monoclonal antibody that targets CD20-positive B cells. While rituximab is an anti-B-cell therapy, it has indirect effects on T cells. B cells often act as "antigen-presenting cells" that activate T cells. By depleting B cells, rituximab deprives T cells of this activating signal, thereby attenuating an overactive T-cell response in autoimmune conditions. It is a safe and effective treatment for many patients with AE[5-7]. In our patient, we administered rituximab after IVIg and steroid therapy because the seizures were refractory. In a review in 2025, rituximab was shown to have a favorable prognosis rate of 80% in cases of AE that did not respond to first-line treatments, especially in patients under 18 years of age or with disease duration ≤180 days[8]. After this treatment, not only seizure resolution but also an improvement in the general clinical condition of our patient were seen.

Although CSF GAD65 antibody testing could not be performed because of financial constraints, the clinical features were highly suggestive of AE. In conjunction with serum GAD65 antibody positivity and a favorable response to immunotherapy, these findings strongly supported a diagnosis of probable GAD65-associated AE.

Our patient had a pre-existing diagnosis of FCD and subsequently developed AE during follow-up. Although chronic neuroinflammatory changes and blood-brain barrier disruption have been reported in FCD, a causal relationship between these findings and the development of AE has not been established. The possible interaction between FCD-associated neuroinflammation and autoimmune processes remains speculative and represents an emerging area of research[9].

Before rituximab treatment, the patient should be free of active or serious infections. A comprehensive evaluation by the immunology and infection departments is recommended. This evaluation includes tests such as a complete blood count, a lymphocyte panel, Ig levels, detailed serology, and a chest X-ray. Post-treatment monitoring is conducted jointly with the immunology department. During this period, attention should be paid to the risk of infection, and live vaccines should be avoided during the first year[10].

Seizure control in anti-GAD65-positive AE cases can be challenging. Rituximab should be considered a safe and effective treatment for patients with AE who have SE that is resistant to first-line treatment.

Footnotes

Authorship Contributions

Surgical and Medical Practices: G.D., Ö.Ö.M., S.S.K., Y.G., F.B., G.S.U., Concept: G.D., Ö.Ö.M., S.S.K., Y.G., F.B., G.S.U., Design: G.D., Ö.Ö.M., S.S.K., Y.G., F.B., G.S.U., Data Collection or Processing: G.D., Ö.Ö.M., S.S.K., Y.G., F.B., G.S.U., Analysis or Interpretation: G.D., Ö.Ö.M., S.S.K., Y.G., F.B., G.S.U., Literature

Search: G.D., Ö.Ö.M., S.S.K., Y.G., F.B., G.S.U., Writing: G.D., Ö.Ö.M., S.S.K., Y.G., F.B., G.S.U.

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

One of the authors of this article (G.S.U.) is a member of the Editorial Board of this journal. She 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.

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

REFERENCES

1. Cellucci T, Van Mater H, Graus F, et al. Clinical approach to the diagnosis of autoimmune encephalitis in the pediatric patient. Neurol Neuroimmunol Neuroinflamm. 2020;7:e663.
2. Yang M, Lian Y. Clinical features and early recognition of 242 cases of autoimmune encephalitis. Front Neurol. 2022;12:803752.
3. Muñoz-Lopetegi A, de Bruijn MAAM, Boukhrissi S, et al. Neurologic syndromes related to anti-GAD65: clinical and serologic response to treatment. Neurol Neuroimmunol Neuroinflamm. 2020;7:e696.
4. Roberti R, Riva A, D'Onofrio G, et al. A proposal for a shared therapeutic algorithm in children with prolonged convulsive seizures and status epilepticus. Expert Rev Neurother. 2024;24:133-8.
5. Liossis SN, Sfikakis PP. Rituximab-induced B cell depletion in autoimmune diseases: potential effects on T cells. Clin Immunol. 2008;127:280-5.
6. Thaler FS, Zimmermann L, Kammermeier S, et al. Rituximab treatment and long-term outcome of patients with autoimmune encephalitis: real-world evidence from the GENERATE registry. Neurol Neuroimmunol Neuroinflamm. 2021;8:e1088.
7. Zuliani L, Nosadini M, Gastaldi M, et al. Management of antibody-mediated autoimmune encephalitis in adults and children: literature review and consensus-based practical recommendations. Neurol Sci. 2019;40:2017-30.
8. Zhang LM, Xing XL, Zhang BR, et al. Efficacy of rituximab as second-line therapy for autoimmune encephalitis: a systematic review and meta-analysis. Heliyon. 2025;11:e41747.
9. Han J, Wang Y, Wei P, Lu D, Shan Y. Unveiling the hidden connection: the blood-brain barrier's role in epilepsy. Front Neurol. 2024;15:1413023.
10. Levin AS, Otani IM, Lax T, Hochberg E, Banerji A. Reactions to rituximab in an outpatient infusion center: a 5-year review. J Allergy Clin Immunol Pract. 2017;5:107-13.e1.

DOI: 10.4274/anatoljmed.2025.63835
Anatol J Gen Med Res 2026;36(2):104-8

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Pulse Wave Analysis in Kidney Transplant Patients Using Tacrolimus

Takrolimus Kullanan Böbrek Nakli Hastalarında Nabız Dalga Analizi

Banu Yılmaz[1], Sibel Akkurt[2], Mehmet Tanrısev[3]

[1]University of Health Sciences Türkiye, İzmir City Hospital, Department of Nephrology, İzmir, Türkiye
[2]University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Internal Medicine, İzmir, Türkiye
[3]University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Nephrology, İzmir, Türkiye

Cite as: Yılmaz B, Akkurt S, Tanrısev M. Pulse wave analysis in kidney transplant patients using tacrolimus. Anatol J Gen Med Res. 2026;36(2):104-8

Abstract

Objective: Pulse wave velocity (PWV) allows the evaluation of arterial stiffness. Although PWV improves after transplantation, it is still higher than in healthy controls. Immunosuppressive agents may contribute to this. Our study aimed to investigate the effect of tacrolimus (TAC) blood level on PWV.

Methods: Fifty-nine kidney transplant patients using TAC with follow-up at the Organ Transplantation Clinic of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, were included in the study. Pulse wave analysis was measured in two measurements using the IEM ambulatory blood pressure measuring device. The relationship between arterial stiffness parameters and the patients' transplantation time, dialysis duration before transplantation, age, kidney functions, proteinuria, and TAC blood levels were examined.

Results: Between PWV and systolic (r=0.589, p<0.001), diastolic (r=0.271, p=0.038) and mean blood pressure (r=0.440, p<0.001) and age (r=0.906, p<0.001) significant correlation was detected. No correlation was detected between TAC level and PWV. When the estimated value for PWV was taken as 8.1 m/s, a significant difference was observed between PWV and age, glomerular filtration rate, Parathormone and proteinuria levels. With TAC level above 7 ng/mL, systolic (120±14 mmHg vs. 119±22 mmHg, p=0.026) and diastolic blood pressure (89±13 mmHg vs. 40±12 mmHg, p=0.023) were found to be higher. A reverse correlation was detected between TAC level and proteinuria (r=-0.316, p=0.035). There was no correlation between transplantation time, pre-transplantation dialysis duration and PWV and other arterial stiffness parameters.

Conclusion: Increase in TAC blood level may cause hypertension in kidney transplant patients. But it is not associated with PWV. PWV is affected only by age, systolic, diastolic and mean arterial pressure in kidney transplant patients using TAC.

Keywords: Tacrolimus, pulse wave velocity, hypertension

Öz

Amaç: Nabız dalga hızı (PWV), arteriyel sertliğin değerlendirilmesini sağlar. Nakil sonrası PWV iyileşmesine rağmen, sağlıklı kontrollere göre halen yüksektir. İmmünosüpresif ajanlar buna katkıda bulunuyor olabilir. Çalışmamızda takrolimus (TAC) kan düzeyinin PWV üzerine etkisini araştırmak amaçlandı.

Yöntem: Sağlık Bilimleri Üniversitesi İzmir Tepecik Eğitim ve Araştırma Hastanesi, Organ Nakli Polikliniği'nde takipli TAC kullanımı olan 59 böbrek nakli hastası çalışmaya dahil edildi. Hastalara IEM ambulatuvar kan basıncı ölçüm cihazı kullanılarak, iki ölçüm ile nabız dalga analizi ölçümü yapıldı. Arteriyel

Address for Correspondence/Yazışma Adresi: Banu Yılmaz, MD, University of Health Sciences Türkiye, İzmir City Hospital, Department of Nephrology, İzmir, Türkiye
E-mail: banuyilmaz16@hotmail.com
ORCID ID: orcid.org/0000-0001-8736-9916

Received/Geliş tarihi: 05.01.2025
Accepted/Kabul tarihi: 02.05.2025
Epub: 03.06.2025
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

sertlik parametreleri ile hastaların nakil süresi, nakilden önceki diyaliz süreleri, yaş, böbrek fonksiyonları, proteinüri, TAC kan düzeyleri arasındaki ilişki incelendi.

Bulgular: PWV ile sistolik (r=0,589, p<0,001), diyastolik (r=0,271, p=0,038) ve ortalama kan basıncı (r=0,440, p<0,001), yaş (r=0,906, p<0,001) arasında anlamlı korelasyon saptanmıştır. TAC düzeyi ile PWV arasında korelasyon saptanmadı. PWV için 8,1 m/s olarak kestirim değeri alındığında PWV ile yaş, glomerüler filtrasyon hızı, parathormon ve proteinüri düzeyleri arasında anlamlı farklılık görüldü. TAC düzeyi 7 ng/mL üzerinde sistolik (120±14 mmHg'ye karşı 119±22 mmHg, p=0,026) ve diyastolik tansiyon (89±13 mmHg'ye karşı 40±12 mmHg, p=0,023) daha yüksek saptandı. TAC düzeyi ile proteinüri arasında ters korelasyon tespit edildi (r=-0,316, p=0,035). Nakil süresi, nakil öncesi diyaliz süresi ile PWV ve diğer arteriyel sertlik parametreleri arasında ilişkili bulunmadı.

Sonuç: TAC kan düzeyindeki artış böbrek nakli hastalarında hipertansiyona neden olabilmektedir. Fakat PWV ile ilişkili değildir. PWV, TAC kullanımı olan böbrek nakli hastalarında sadece yaş, sistolik, diyastolik ve ortalama arteriyel basınçtan etkilenmektedir.

Anahtar Kelimeler: Takrolimus, nabız dalga analizi, hipertansiyon

Introduction

Arteriosclerosis and its complications are the leading causes of death in renal transplant recipients[1]. Pulse wave velocity (PWV) is a sensitive parameter that allows the regional assessment of arterial stiffness. PWV is an independent determinant of arterial stiffness and increased cardiovascular risk in patients with hypertension and dialysis[2,3]. Despite improvements in PWV after kidney transplantation, it is still higher than in healthy controls[4,5]. Post-transplant immunosuppressive regimens, especially cyclosporine-containing regimens, seem to diminish the beneficial vascular effects of kidney transplantation. When the effects of calcineurin inhibitors (CNIs) [CNI tacrolimus (TAC) and cyclosporine-A (CsA)] on arterial stiffness are compared, CsA is known to increase arterial stiffness compared to TAC[6-8]. In a previous study conducted on patients using TAC, no significant relationship was found between TAC use and PWV values, though the study did find a relationship between PWV and age[7]. However, a relationship was shown between CsA level and PWV[9]. The aim of our study was to examine the relationship between TAC level and PWV in patients who adopted the TAC regimen after transplantation. In addition, the relationship between pre-transplant dialysis duration, post-transplant duration, the renal functions, and PWV was examined.

Materials and Methods

Fifty-nine consecutive kidney transplant patients using TAC who were followed up at the University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital Organ Transplantation Outpatient Clinic between 2022-2023 were included in the study. The study was approved by the Non-Interventional Ethics Committee of the University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital (approval no: 2024/09-11, date: 10.10.2024). Kidney transplant recipients with a transplant duration under one year, patients with atrial fibrillation, and patients with glomerular filtration rate (GFR) below 30 mL/min/m^2, were not included in the study. The study was designed as a retrospective case-control study. In routine outpatient clinic controls of 59 patients, pulse wave analysis was measured twice using an IEM ambulatory blood pressure measurement device. PWV and other arterial stiffness parameters [augmentation index (AIx) and pressure, pulse pressure, peripheral resistance], systolic blood pressure (SBP), diastolic blood pressure (DBP) and mean blood pressures (MBPs) were recorded with the device. Kidney transplantation dates; hemodialysis or peritoneal dialysis durations; TAC usage durations; additional diseases (diabetes, cardiovascular disease); smoking; medications used; creatinine on the day of pulse wave analysis; GFR; mean of the last 3 TAC levels of the patients; uric acid; calcium; phosphorus; parathyroid hormone (PTH); 25 (OH) vitamin D; spot urine protein creatinine ratios were recorded retrospectively from the outpatient clinic files. The relationship between the parameters measured by pulse wave analysis and TAC level, transplantation time, dialysis duration before transplantation, creatinine, GFR, uric acid, PTH, vitamin D, calcium, and phosphorus levels was examined.

Statistical Analysis

SPSS version 27.0 (IBM Corp., Armonk, NY, USA) was used in the analysis of the data. The distribution of the data was analyzed with Kolmogorov-Smirnov and Shapiro-Wilk tests.

In the comparison of means, Student's t-test was used for those showing normal distribution and Mann-Whitney U test was used for those not showing normal distribution. The chi-square test was used in the examination of categorical data. In the examination of the correlation of continuous data, Spearman and Pearson correlation tests were applied according to the distribution of the data. P<0.05 was accepted as statistically significant.

Results

The baseline demographic data of the patients are presented in Table 1. Thirty-three female and twenty-six male patients were included in the study. A significant correlation was found between PWV and age (r=0.906, p<0.001), SBP (r=0.589, p<0.001), DBP (r=0.271, p=0.038), and MBP (r=0.440, p<0.001). No correlation was found between TAC level and PWV. When the cut-off value for PWV was taken as 8.1 m/s, a significant difference was found among PWV, age, GFR, PTH, and proteinuria levels (Table 2).

When patients with TAC levels above and below 7 ng/mL were compared in terms of arterial stiffness parameters, no relationship was found between the parameters and TAC levels. SBP (120.6±14.6 mmHg vs. 119.7±22.1 mmHg, p=0.026) and DBP (89±13 mmHg vs. possibly incorrect value 40±12 mmHg, p=0.023) were higher with TAC levels above 7 ng/mL, compared to lower TAC levels. A negative correlation was found between TAC levels and proteinuria (r=-0.316, p=0.035). No relationship was found between transplantation duration, pre-transplantation dialysis duration, and PWV and other arterial stiffness parameters.

Table 1. Baseline demographic data

Variables	Mean ± SD	n
Age (years)	45.8±12.9	59
Kidney transplantation duration (months)	122.2±79.6	59
Pre-transplant dialysis duration (months)	26.2±34.1	59
Creatinine (mg/dL)	1.41±0.49	59
GFR (mL/min)	60.3±22.7	59
Calcium (mg/dL)	9.6±0.5	59
Phosphorus (mg/dL)	3.4±0.7	59
Uric acid (mg/dL)	6.7±1.5	59
PTH (pg/mL)	150±91.5	59
Vitamin D (ng/mL)	16.7±7.5	57
Hemoglobin (g/dL)	12.9±1.9	59
Spot urine protein creatinine ratio (mg/mg)	0.577±1.024	59
MBP (mmHg)	105.6±13.6	59
Systolic BP (mmHg)	120±20	59
Diastolic BP (mmHg)	78.8±24.7	59
PWV m/sec	7.01±1.49	59

MBP: Mean blood pressure, BP: Blood pressure, SD: Standard deviation, PTH: Parathyroid hormone, GFR: Glomerular filtration rate, PWV: Pulse wave velocity

Discussion

PWV is a valuable parameter for assessing arterial stiffness. Successful kidney transplantation leads to a decrease in PWV[4,10]. However, PWV is still higher than in healthy controls[11]. This situation seems to be related to the immunosuppressive protocol. In studies comparing arterial stiffness parameters in CsA and TAC-containing regimens, AIx and PWV were found to be higher in the CsA-containing regimen[4,7]. A significant relationship was found between TAC and age, but no data were obtained indicating that TAC increased PWV[7]. In our study, no significant correlation was found between TAC level and PWV.

There is a dose-response relationship between increasing PWV and the occurrence of major cardiovascular outcomes such as acute coronary syndrome, cerebrovascular syndromes, and death. Regardless of the threshold value, cardiovascular risk increases[12]. In a prospective study conducted in a cohort of 512 kidney transplant recipients in 2011, cardiovascular survival was found to be worse in patients with PWV ≥8.1 m/s compared to patients with PWV <8.1 m/s[13]. Therefore, the cut-off value for PWV in our study was taken as 8.1 m/s. When PWV was measured at 8.1 m/s, a significant relationship was found between PWV and GFR, PTH, proteinuria, and age. In patients with higher PWV, age was higher, GFR was lower, and PTH and proteinuria were increased. Classical risk factors influencing PWV include hypertension, age, PTH, and proteinuria[14-16]. Despite being a CNI, the lack of association between TAC levels and PWV in our study may be attributed to the narrow

Table 2. Patients with PWV above and below 8.1 m/sec

	Group 1, n=21	Group 2, n=38	
PWV (m/sec)	>8.1 m/sec	<8.1 m/sec	p-value
GFR (mL/min)	53.07±16.88	62.58±24.04	0.044[b]
Spot urine protein creatinine ratio (mg/mg)	1.171±1.304	0.408±0.880	0.036[b]
Age (years)	60.9±6.3	41.1±10.6	0.011[a]
PTH (pg/mL)	175.3±136.9	141.8±72.2	0.013[a]

[a]: Mann-Whitney U test, [b]: Student t-test, PWV: Pulse wave velocity, PTH: Parathyroid hormone, GFR: Glomerular filtration rate

therapeutic range of TAC, as well as the prevention of sustained high levels over prolonged periods.

Age and blood pressure are the main factors affecting the PWV[17]. In our study, a significant relationship between PWV and SBP, DBP, MBP, and age was found, in line with previous studies. No relationship was found between PWV and transplantation duration, pre-transplantation dialysis duration, and renal functions.

The cause of post-transplant hypertension is multifactorial, and one of the risk factors is CNI[14]. TAC-based regimens may be associated with a lower incidence of hypertension compared to CsA-based regimens[18,19]. In our study, no correlation was found between hypertension and TAC levels in the entire patient group, but increased systolic and DBP was detected in the patient group with TAC levels of 7 ng/mL. This showed us that increased TAC levels may be associated with hypertension. While it is well-established that TAC can lead to increased blood pressure (both SBP and DBP), our study did not observe an impact on PWV despite the hypertensive effects of TAC. This discrepancy could be attributed to several factors, including PWV may be influenced by other mechanisms independent of blood pressure, or that the duration of hypertension caused by TAC in our cohort was insufficient to lead to measurable changes in arterial stiffness. Arterial stiffness can be a consequence of long-term hypertension, but initially, hypertension may develop without causing significant arterial stiffening. In other words, it is possible for hypertension to develop before it leads to an increase in PWV[20]. Further investigation into the complex relationship between TAC, hypertension, and arterial stiffness is warranted to better understand these dynamics and identify other contributing factors.

Study Limitations

Our study's limitations were due to the lack of molecular markers that could predict arterial stiffness before PWV values were available. In addition, no comparison was made with other agents used in kidney transplantation, such as CsA, sirolimus, and everolimus. Only patients using TAC were included in the study.

Conclusion

In our study, the only factors affecting PWV were determined to be blood pressure and age. Although TAC levels above 7 ng/mL were found to be associated with hypertension, it was observed that they did not affect arterial stiffness parameters.

Ethics

Ethics Committee Approval: The study was approved by the Non-interventional Ethics Committee of the University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital (approval no: 2024/09-11, date: 10.10.2024).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

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

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

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

References

1. Kasiske BL. Epidemiology of cardiovascular disease after renal transplantation. Transplantation. 2001;72(Suppl 6):S5-8.
2. Blacher J, Asmar R, Djane S, London GM, Safar ME. Aortic pulse wave velocity as a marker of cardovascular risk in hypertensive patients. Hypertension. 1999;33:1111-7.
3. Blacher J, Guerin AP, Pannier B, Marchais SJ, Safar ME, London GM. Impact of aortic stiffness on survival in endstage renal disease. Circulation. 1999;99:2434-9.
4. Zaungas S, Kerr PG, Chadban S, et al. Arterial function after successful renal transplantation. Kidney Int. 2004;65:1882-9.
5. Cohen DL, Warburton KM, Warren G, Bloom RD, Townsend RR. Pulse wave analysis to assess vascular compliance changes in stable renal transplant recipients. Am J Hypertens. 2004;17:209-12.
6. Ferro CT, Savage T, Pinder SJ, Tomson CR. Central aortic pressure augmentation in stable renal transplant recipients. Kidney Int. 2002; 62:166-71.
7. Strózecki P, Adamowicz A, Włodarczyk Z, Manitius J. The influence of calcineurin inhibitors on pulse wave velocity in renal transplant recipients. Ren Fail. 2007;29:679-84.
8. Seibert F, Behrendt C, Schmidt S, van der Giet M, Zidek W, Westhoff TH. Differential effects of cyclosporine and tacrolimus on arterial function. Transpl Int. 2011;24:708-15.
9. Sugianto RI, Ostendorf K, Bauer E, et al. Arterial stiffness and blood pressure increase in pediatric kidney transplant recipients. Pediatr Nephrol. 2023;38:1319-27.
10. Covic A, Goldsmith DJ, Gusbeth-Tatomir P, Buhaescu I, Covic M. Successful renal transplantation decreases aortic stiffness and increases vascular reactivity in dialysis patients. Transplantation. 2003;76:1573-7.
11. Bahous SA, Stephan A, Barakat W, Blacher J, Asmar R, Safar ME. Aortic pulse wave velocity in renal transplant patients. Kidney Int. 2004;66:1486-92.

12. Williams B, Mancia G, Spiering W, et al. 2018 ESC/ESH guidelines for the management of arterial hypertension: the task force for the management of arterial hypertension of the European society of cardiology and the European society of hypertension. J Hypertens. 2018;36:1953-2041.
13. Verbeke F, Maréchal C, Van Laecke S, et al. Aortic stiffness and central wave reflections predict outcome in renal transplant recipients. Hypertension. 2011;58:833-8.
14. Marshall AG, Neikirk K, Afolabi J, et al. Update on the use of pulse wave velocity to measure age-related vascular changes. Curr Hypertens Rep. 2024;26:131-40.
15. Zhu H, Xue H, Wang G, Fu Z, Liu J, Shi Y. [Association between urinary microalbumin-to-creatinine ratio and brachial-ankle pulse wave velocity in hypertensive patients]. Zhonghua Xin Xue Guan Bing Za Zhi. 2015;43:308-11. Chinese.
16. Yeung WG, Toussaint ND, Lioufas N, et al. Vitamin D status and intermediate vascular and bone outcomes in chronic kidney disease: a secondary post hoc analysis of IMPROVE-CKD. Intern Med J. 2024;54:1960-9.
17. Hoorn EJ, Walsh SB, McCormick JA, et al. The calcineurin inhibitor tacrolimus activates the renal sodium chloride cotransporter to cause hypertension. Nat Med. 2011;17:1304-9.
18. Artz MA, Boots JM, Ligtenberg G, et al. Conversion from cyclosporine to tacrolimus improves quality-of-life indices, renal graft function and cardiovascular risk profile. Am J Transplant. 2004;4:937-45.
19. Weir MR, Burgess ED, Cooper JE, et al. Assessment and management of hypertension in transplant patients. J Am Soc Nephrol. 2015;26:1248-60.
20. Skov J, Nedergaard S, Andreasen M. New type of synaptically mediated epileptiform activity independent of known glutamate and GABA receptors. J Neurophysiol. 2005;93:1845-56.

DOI: 10.4274/anatoljmed.2025.57070
Anatol J Gen Med Res 2026;36(2):109-17

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Prognostic Factors for 30-day Mortality after Hip Fracture in the Elderly

Yaşlılarda Kalça Kırığı Sonrası 30 Günlük Mortalitede Prognostik Faktörler

Büşra Akpınar[1], Mürsel Karadavut[1], Eyüp Şenocak[2], İbrahim Dağ[2], Ömer Karaşahin[3], Pınar Tosun Taşar[1]

[1]Atatürk University Faculty of Medicine, Department of Internal Medicine, Division of Geriatrics, Erzurum, Türkiye
[2]Atatürk University Faculty of Medicine, Department of Orthopedic Surgery, Erzurum, Türkiye
[3]Erzurum Training and Research Hospital, Clinic of Infectious Diseases and Clinical Microbiology, Erzurum, Türkiye

Cite as: Akpınar B, Karadavut M, Şenocak E, Dağ İ, Karaşahin Ö, Tosun Taşar P. Prognostic factors for 30-day mortality after hip fracture in the elderly. Anatol J Gen Med Res. 2026;36(2):109-17

Abstract

Objective: In older adults, hip fractures lead to disability, diminished life quality, and increased socioeconomic burden and risk of death. This study aimed to evaluate the prognostic determinants of survival within a geriatric cohort following hip fracture.

Methods: We conducted a retrospective chart review of individuals aged 60 and above admitted with hip fractures to a tertiary academic orthopedic department from January 2013 through January 2024. Our analysis integrated baseline demographic and clinical profiles, fracture etiology, and anatomical site. Furthermore, we assessed surgical variables (timing and specific techniques employed), postoperative adverse events, total hospitalization duration, and mortality status at one month post-injury.

Results: A total of 250 older adults (median age: 80 years, range: 60-110, 62.4% women) constituted the study sample. The 30-day mortality rate was 6.8% (n=17). Patients in the non-surviving group were notably older (p=0.021) and more likely to have dementia (p=0.014) than those who survived. Femoral neck fractures were associated with improved survival (p=0.004), whereas pertrochanteric fractures were associated with increased mortality (p<0.001). Admission laboratory data showed that non-survivors had depressed lymphocyte counts and elevated serum creatinine and blood urea nitrogen concentrations. Subsequent multivariate logistic regression identified pre-existing dementia (odds ratio=4.12) and pertrochanteric fracture location (odds ratio=15.87) as independent predictors of early mortality.

Conclusion: Early mortality following a hip fracture among geriatric patients was strongly predicted by the presence of comorbid dementia, pertrochanteric fracture, and admission lymphopenia. Recognizing these baseline clinical vulnerabilities is critical for implementing vigilant, targeted perioperative care protocols.

Keywords: Elderly, hip fracture, mortality, prognostic factors, risk factors

Address for Correspondence/Yazışma Adresi: Mürsel Karadavut, MD, Atatürk University Faculty of Medicine, Department of Internal Medicine, Division of Geriatrics, Erzurum, Türkiye
E-mail: mursel_karadavut@hotmail.com
ORCID ID: orcid.org/0000-0002-0099-6317

Received/Geliş tarihi: 03.02.2025
Accepted/Kabul tarihi: 25.10.2025
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

Amaç: Kalça kırıkları yaşlılarda ciddi sakatlığa yol açmakta, yaşam kalitesinde düşüşe sebep olmakta, mortalite riskini ve sosyoekonomik yükü artırmaktadır. Çalışmamızda kalça kırığı olan geriatrik hastalarda mortalite ile ilişkili faktörleri belirlemeyi amaçladık.

Yöntem: Ocak 2013 ile Ocak 2024 tarihleri arasında 3. basamak bir üniversite hastanesi olan hastanemizin ortopedi kliniğinde kalça kırığı nedeniyle yatan geriatrik hastaların verileri geriye dönük olarak incelendi. Demografik ve klinik verilerin yanı sıra kalça kırığının yeri ve nedeni, preoperatif bekleme süresi, operasyon türü, postoperatif komplikasyonlar, hastanede kalış süresi ve 30 günlük sağ kalım bilgileri kaydedildi.

Bulgular: Çalışmaya 60 yaş ve üzeri 250 hasta dahil edildi. Ortanca yaş 80 (60-110) yıldı ve 156'sı (%62,4) kadındı. On yedi hastanın (%6,8) travma sonrası kalça kırığı nedeniyle 30 gün içinde öldüğü görüldü. Otuz gün içerisinde mortalite gelişen hastalar istatistiksel açıdan anlamlı olarak daha yaşlıydı ve daha yüksek bir demans prevalansına sahipti (sırasıyla p=0,021 ve p=0,014). Femur boyun kırığı olanlar hastalarda mortalite daha az iken (p=0,004) pertrokanterik kırığı olanlarda ise mortalite daha fazla idi (p<0,001). Kan üre azotu ve kreatinin değeri ölenlerde anlamlı olarak daha yüksekken, lenfosit sayısı anlamlı olarak daha düşüktü. Çok değişkenli lojistik regresyon analizinde demans varlığının 4,12 kat, pertorkanterik kırık varlığının 15,87 kat 30 günlük mortalite riskini artıran bağımsız risk faktörü olduğu saptandı.

Sonuç: Yaşlılarda kalça kırığının mortalitesi demans, pertorkanterik kırıklar ve lenfopenisi olan hastalarda daha yüksekti. Bu hastalar yakından takip edilmelidir.

Anahtar Kelimeler: Yaşlı, kalça kırığı, mortalite, prognostik faktörler, risk faktörleri

Introduction

The prevalence of osteoporosis rises with advancing age due to a combination of altered bone microarchitecture, diminished osteoblastic function, and heightened osteoclastic activity. Furthermore, advanced age is associated with an increased risk of falls secondary to visual and balance impairments, frailty, polypharmacy, gait instability, and sarcopenia. Consequently, osteoporotic individuals are at increased risk of hip fractures[1], which are among the most important orthopedic challenges in the geriatric population[2]. Beyond increasing mortality risk and socioeconomic burden, these fractures severely compromise quality of life and result in serious disability. With the ongoing aging of the global population, projections indicate a continuous climb in the yearly incidence of these injuries[3,4], with estimates anticipating up to 6.25 million cases by the year 2050[5].

Within the first 12 months following a hip fracture, the observed fatality rate in the geriatric population ranges from 14% to 36%[6]. However, evidence suggests that three-quarters of hip fracture-related deaths are associated with preexisting comorbidities rather than the fracture itself[7]. Additional variables shown to elevate post-fracture mortality risk include male sex, older age, institutionalization (such as nursing home residency), malnutrition elevated American Society of Anesthesiologists (ASA) classification, and specific conditions such as diabetes, dementia, cancer, and ischemic heart disease[8-10]. Typically, clinical guidelines recommend operative management of these fractures in elderly patients within 48 hours. While prompt surgical intervention is generally favored to mitigate both morbidity and mortality in the majority of cases, the effect of surgical timing on mortality is debated in the literature[11-13].

While the broader literature contains numerous epidemiological investigations of hip fractures, there is a notable scarcity of data specifically addressing post-fracture mortality among older adults in the national context. The primary objective of this study was to identify prognostic factors associated with survival in a cohort of older patients with hip fracture.

Materials and Methods

This cross-sectional, retrospective study included geriatric patients (age 60 and above) who were admitted with hip fractures to the orthopedics department of a tertiary academic medical hospital between January 2013 and January 2024. Patients who met the above criteria and received a radiological diagnosis (via computed tomography and/or direct X-ray) of femoral neck, pertrochanteric, or subtrochanteric fractures were eligible for inclusion. Patients were excluded if they refused hospitalization, presented with a pathological fracture, or had a documented prior hip fracture.

Patient files and electronic databases were reviewed to extract a comprehensive set of variables. These included survival status at 30 days, duration of inpatient stay, occurrence of postoperative adverse events, need for perioperative blood product transfusion, estimated blood loss, anesthesia modality (general, spinal, or epidural), ASA

classification, surgical technique, date of the fracture and subsequent surgery, time from injury to operation, presence of concomitant non-hip fractures, anatomical location and etiology of the hip injury, osteoporosis subtype and bone mineral density metrics, pharmacological profile (type and quantity of medications), existing chronic comorbidities, and baseline demographic details. Mortality outcomes were verified through the national death registry, and postoperative survival times were recorded.

Osteoporotic conditions were grouped into primary and secondary. Traumatic etiology was dichotomized into high-energy and low-energy mechanisms. We stratified postoperative adverse events into surgical and metabolic categories. Specifically, surgical complications encompassed surgical site infections, hip dislocations, and stress fractures, whereas metabolic complications included episodes of delirium, acute kidney injuries, cerebrovascular accidents, pulmonary embolisms, pneumonias, and respiratory failures. To quantify the cumulative impact of concurrent illnesses, we utilized the modified Charlson comorbidity index (mCCI). The mCCI was developed by Charlson et al.[14] in 1987 to estimate the 10-year mortality risk by measuring the burden of comorbid disease.

Additionally, to assess potential associations with mortality, a wide array of admission laboratory parameters was analyzed. This panel comprised levels of gamma-glutamyl transferase, alkaline phosphatase, magnesium, phosphorus, calcium, potassium, glucose, sodium, blood urea nitrogen (BUN), albumin, total protein, lactate dehydrogenase, alanine transaminase, aspartate transaminase, hematocrit, and hemoglobin, and counts of platelets, lymphocytes, neutrophils, and total white blood cells.

Prior to the commencement of the study, formal approval was obtained from the Clinical Research Ethics Committee of Atatürk University Faculty of Medicine (approval no: 06, date: 21.02.2024).

Statistical Analysis

Data analysis was conducted in IBM SPSS Statistics version 21.0 (IBM Corp). The normality of distributions of continuous variables was assessed using the Kolmogorov-Smirnov test. For parametric data, the Student's t-test was used to perform pairwise comparisons between independent groups; for non-parametric data, the Mann-Whitney U test was used. Associations among categorical variables were examined using chi-square tests. To evaluate predictors of early mortality, parameters that achieved statistical significance in univariate analyses were included in a multivariate logistic regression analysis (backward LR model with entry and removal criteria set at 0.05 and 0.10, respectively). The threshold for statistical significance was defined as a p-value below 0.05.

Results

A total of 250 older adults (aged 60 years and older) with hip fractures were analyzed. Women constituted 62.4% (n=156) of the cohort, and the median age was 80 years (range: 60-110). The 30-day mortality rate was 6.8% (n=17). Surgical intervention was the primary treatment modality for the majority of patients (97.2%, n=243).

Table 1 details the distribution of baseline demographics, comorbidities, and hip fracture characteristics, stratified by early mortality status. Compared with surviving patients, non-survivors were older (p=0.021) and more likely to have pre-existing dementia (p=0.014). Furthermore, the non-surviving group exhibited a higher median mCCI score (p=0.030). Anatomical fracture site was also significantly associated with 30-day survival outcomes, with femoral neck fractures associated with better survival (p=0.004) and pertrochanteric fractures associated with higher mortality (p<0.001). Individuals managed conservatively (without surgery) exhibited a higher rate of early mortality than those who underwent operative repair (p=0.008).

Table 2 outlines the relationship between 30-day mortality and perioperative factors, including the need for red-cell transfusion, anesthetic modality, time to surgery, and postoperative adverse events. Overall, 54 patients (22.2%) experienced at least one postoperative medical or surgical complication; of these, 7 developed multiple complications. The rate of early mortality was higher in patients who experienced postoperative adverse events (64.3% vs. 20.1%; p<0.001). Figure 1 illustrates the frequency of complications according to 30-day survival status. The most common surgical complication was surgical site infection (n=12), while the most common medical complications were pulmonary embolism (n=15) and acute kidney injury (n=16).

Table 3 summarizes the association between 30-day mortality and admission laboratory results. Non-survivors had significantly lower lymphocyte levels (p=0.003) and higher serum concentrations of creatinine (p=0.030) and BUN (p=0.006).

Table 1. Distribution of demographic characteristics, comorbidities, and features of hip fracture according to 30-day mortality			
	30-day mortality		p-value
	Yes (n=17)	No (n=233)	
Age, median (IQR)	83.0 (80.5-86.5)	79.0 (72.0-84.0)	**0.021**
Sex (female), n (%)	11 (64.7)	145 (62.2)	0.839
Comorbidity, n (%)			
HT	8 (47.1)	134 (57.5)	0.401
DM	2 (11.8)	40 (17.2)	0.431
CAD	6 (35.3)	76 (32.6)	0.821
CHF	2 (11.8)	17 (7.3)	0.377
CKD	6 (35.3)	39 (16.7)	0.055
CVD	3 (17.6)	27 (11.6)	0.458
HL	1 (5.9)	6 (2.6)	0.393
Malignancy	2 (11.8)	21 (9.0)	0.478
Thyroid disease	3 (17.6)	20 (8.6)	0.197
Dementia	8 (47.1)	49 (21.0)	**0.014**
Depression	-	17 (7.3)	0.290
Parkinson's	1 (5.9)	7 (3.0)	0.435
COPD	1 (5.9)	29 (12.4)	0.369
Asthma	1 (5.9)	5 (2.1)	0.348
Number of diseases, median (IQR)	3 (1-6)	3 (1-4)	0.436
mCCI score, median (IQR)	5 (4-6)	4 (4-6)	**0.030**
Number of medications used, median (IQR)	2 (0-5)	4 (1-6)	0.192
Cause of hip fracture, n (%)			0.488
Low-energy trauma	17 (100.0)	223 (95.7)	
High-energy trauma	-	10 (4.3)	
Location of hip fracture, n (%)			**0.001**
Femoral neck fracture	5 (29.4)	152 (65.2)	
Pertrochanteric fracture	12 (70.6)	64 (27.5)	
Subtrochanteric fracture	-	17 (7.3)	
Treated surgically, n (%)	14 (82.4)	229 (98.3)	**0.008**

IQR: Interquartile range, HT: Hypertension, DM: Diabetes mellitus, CAD: Coronary artery disease, CHF: Congestive heart failure, CKD: Chronic kidney disease, CVD: Cerebrovascular disease, HL: Hyperlipidemia, COPD: Chronic obstructive pulmonary disease, mCCI: Modified Charlson comorbidity index

To identify independent predictors of early mortality, we constructed a multivariate logistic regression model incorporating all variables that reached statistical significance in univariate testing (admission lymphocyte count, serum creatinine, serum BUN, surgical treatment, anatomical fracture site, dementia diagnosis, and patient age). As detailed in Table 4, this analysis identified pertrochanteric fracture location [odds ratio (OR)=15.87, p<0.001] and comorbid dementia (OR=4.12, p=0.021) as primary independent risk factors. Additionally, each single-unit decrease in baseline lymphocyte count correlated with a 0.2% increase in the odds of early mortality (OR=1.002, p=0.004). In contrast, operative treatment emerged as a robust protective factor, yielding a 97.7% reduction in the odds of mortality (OR=0.023, p<0.001).

Discussion

This study evaluated a cohort of 250 geriatric patients with hip fractures (median age 80 years; 62.4% women) to determine risk factors for early mortality. We determined that the odds of 30-day mortality were more than 4 times higher for patients with dementia and nearly 16 times higher for patients with pertrochanteric fracture. A lower

Table 2. Distribution of postoperative complications, time to surgery, type of anesthesia, and red cell transfusion according to 30-day mortality among operated patients (n=243)			
	30-day mortality		
	Yes (n=14)	**No (n=229)**	**p-value**
Surgery performed, n (%)			0.299
Hip replacement	14 (100)	195 (85.2)	
Proximal femoral pin	-	23 (10.0)	
Dynamic hip pin	-	11 (4.8)	
Postoperative complication, n (%)	9 (64.3)	46 (20.1)	**<0.001**
Type of complication, n (%)			0.293
Medical	8 (88.9)	31 (68.9)	
Surgical	-	10 (22.2)	
Medical+surgical	1 (11.1)	4 (8.9)	
Time to surgery (h), n (%)			0.273
0-12	3 (21.4)	14 (6.1)	
12-24	3 (21.4)	42 (18.3)	
24-36	1 (7.1)	27 (11.8)	
36-48	2 (14.3)	40 (17.5)	
>48	5 (14.3)	106 (46.3)	
Anesthesia received, n (%)			0.799
Spinal	6 (42.9)	109 (47.6)	
Epidural	5 (35.7)	86 (37.6)	
General	3 (21.4)	34 (14.8)	
Postoperative red cell transfusion, n (%)	8	143	0.375

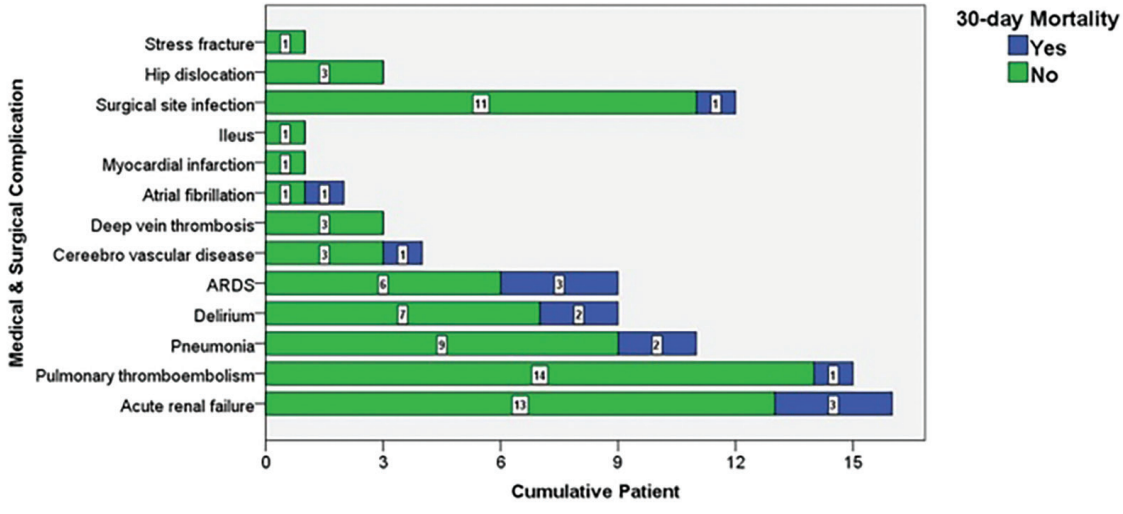


Figure 1. Distribution of medical and surgical complications by 30-day mortality

ARDS: Acute respiratory distress syndrome

Table 3. Comparison of laboratory parameters at hospital admission according to 30-day mortality

Laboratory parameters, median (IQR)	30-day mortality		
	Yes (n=17)	No (n=233)	p-value
Leukocytes (10^3/µL)	10.10 (7.94-12.10)	10.10 (8.06-12.41)	0.854
Lymphocytes (10^3/µL)	0.89 (0.68-1.05)	1.19 (0.86-1.68)	**0.003**
Neutrophils (10^3/µL)	8.34 (6.19-10.41)	7.87 (5.83-1.68)	0.540
Hb (g/dL)	11.50 (9.80-13.75)	12.70 (11.55-13.75)	0.102
Hct (%)	35.90 (30.35-40.40)	38.40 (34.85-41.40)	0.157
PLT (10^3/µL)	212 (154-251)	227 (184-290)	0.113
BUN (mg/dL)	26.17 (22.08-37.62)	21.5 (16.82-28.5)	**0.006**
Creatinine (mg/dL)	1.02 (0.86-1.12)	0.83 (0.63-1.1)	**0.030**
Na (mEq/L)	139 (137-141)	138 (136-140)	0.357
K (mEq/L)	4.09 (3.91-4.34)	4.14 (3.75-4.47)	0.988
AST (IU/L)	21 (16-37)	23 (18-28.8)	0.793
ALT (IU/L)	12 (11-14)	14 (10-19)	0.574
Ca (mg/dL)	8.7 (8.3-9.2)	8.8 (8.405-9.2)	0.541
P (mg/dL)	3.0 (2.4-3.3)	3.04 (2.5-3.5)	0.590
Mg (mg/dL)	1.71 (1.63-2.21)	1.87 (1.7-2.04)	0.490
Albumin (g/dL)	3.47 (3.13-3.91)	3.645 (3.36-3.9325)	0.283
ALP (IU/L)	103 (84-138)	86 (68-109.25)	0.061
Total protein (g/dL)	6.30 (5.95-7.2)	6.85 (6.41-7.30)	0.136

IQR: Interquartile range, Hb: Hemoglobin, Hct: Hematocrit, PLT: Platelet, BUN: Blood urea nitrogen, Na: Sodium, K: Potassium, AST: Aspartate transaminase, ALT: Alanine transaminase, Ca: Calcium, P: Phosphorus, Mg: Magnesium, ALP: Alkaline phosphatase

Table 4. Results of multivariate logistic regression of risk factors significant for 30-day mortality

Variables	B	OR	95% CI	p-value
Presence of dementia	1.417	4.124	1.242-13.695	**0.021**
Undergoing surgery	-3.760	0.023	0.003-0.172	**<0.001**
Pertrochanteric fracture	2.765	15.874	3.882-64.910	**<0.001**
Lymphocyte count	0.002	1.002	1.001-1.004	**0.004**

OR: Odds ratio, CI: Confidence interval

lymphocyte count was also significant (0.2% lower odds per unit decrease in lymphocyte count). Conversely, surgical treatment reduced the odds of early mortality by 97.7%.

A previous prospective study showed that in older adults, new functional losses occur throughout the first year after a hip fracture[15].Other studies demonstrated that the prevalence of frailty increased with decreased mobility, impaired balance, decreased muscle strength, impaired cognitive function, malnutrition, reduced physical activity, and greater falls risk[16-19]. In a cohort of women aged 65 and over, followed for 12 months post-hip fracture, elevated serum interleukin-6 levels were associated with persistent lower-extremity dysfunction. For patients with multiple pre-existing comorbidities, a hip fracture may trigger or accelerate frailty via inflammatory or immunological mechanisms, which can impact survival[20]. In this study, we aimed to identify independent predictors of mortality risk among older patients in whom hip fracture is known to be associated with a poorer prognosis.

In our geriatric cohort, the observed 30-day mortality rate after hip fracture was 6.8%. This rate ranges between 5.4% and 14.3% in the published literature[21,22]. Furthermore, our analysis identified dementia in 22.8% of the study population. Past research has reported a wide variation in

dementia prevalence (6.38% to 30%) among older adults with hip fracture, depending on variables such as genetic background, ethnicity, and demographic factors. Dementia is a known risk factor for frailty because it contributes to malnutrition, infection, and trauma, and it has also been linked to higher mortality after hip fracture[23,24]. Similarly, a review by Lai et al.[25] including 8.080 patients with dementia and 145.543 cognitively intact individuals confirmed that cognitive decline exacerbates 1- and 2-year mortality following a hip fracture. Furthermore, Bai et al.[26] synthesized 18 cohort studies in a meta-analysis, concluding that comorbid dementia multiplied the risk of mortality at 1, 6, and 12 months by 1.57, 1.97, and 1.77 times, respectively. In the present cohort, dementia was associated with an over 4-fold increase in the odds of 30-day mortality.

Surgical intervention remains the standard approach for managing hip fractures, as reflected by the 97.2% operation rate in our study. Surgical repair is considered crucial for expediting recovery and facilitating rehabilitation. Conversely, older individuals who are managed conservatively face heightened mortality risks associated with extended periods of bed rest. Consequently, our finding that surgery is an independent protective factor against early mortality is consistent with clinical expectations.

Regarding adverse events following surgery, at least one medical or surgical complication occurred in 22.2% of our cohort, and multiple complications affected 2.9% of our cohort. The most common medical complications were acute kidney injury and pulmonary embolism, and the most common surgical complication was surgical site infection. As anticipated, the occurrence of these postoperative complications was associated with higher 30-day mortality than in complication-free patients. In their 2019 study, Saul et al.[27] reported complications in 33% of patients, 11.5% of whom had surgical complications. Other reports have identified electrolyte imbalances and anemia as the prevailing non-surgical issues in this patient population[28].

Investigating surgical modalities, Parker and Gurusamy[29] concluded in a 2006 systematic review of 17 studies that total arthroplasty and internal fixation yielded no significant mortality differences in older adults with intracapsular fractures. Likewise, we observed no difference in mortality with respect to operative technique, although the anatomical location of the fracture substantially influenced outcomes. Mortality rates have been shown to be higher among individuals with subtrochanteric and trochanteric fractures relative to those with femoral neck fractures[30]. Additionally, extracapsular fractures are broadly recognized to portend worse survival outcomes than intracapsular fractures[31]. Ricci et al.[32] also documented elevated mortality rates among patients presenting with subtrochanteric or trochanteric fractures. In line with these findings, our data demonstrated a 15.87-fold increase in the odds of early mortality for patients with pertrochanteric fractures. Moreover, consistent with our own observations, prior research confirms that survival rates are generally more favorable following intracapsular trauma compared to extracapsular events[27].

Within the context of hip fractures, the effect of operative timing on patient survival remains a subject of debate. According to a meta-analysis by Moja et al.[33] delayed intervention correlated with a marked rise in pressure ulcer formation and overall mortality risk. Therefore, the general recommendation is that hip fractures be surgically repaired within a 24- to 48-hour window. Conversely, a 2018 prospective cohort study by Lizaur-Utrilla et al.[34] found no increased risk of mortality or complications when surgery was postponed for two or more days to medically optimize highly comorbid individuals. However, when organizational deficits caused the surgical delay, the researchers noted higher rates of one-year mortality and postoperative adverse events. Within our cohort, the interval between the traumatic event and the operation did not significantly alter 30-day survival outcomes. Given that our institution prioritizes prompt surgical fixation whenever feasible, we hypothesize that any operative delays in our cohort were deliberate measures to achieve adequate medical stabilization prior to surgery.

Another finding of our study was that patients who experienced early mortality following traumatic hip fracture had higher BUN and creatinine values and lower admission lymphocyte counts. Our findings are consistent with the results of previous studies. Blanco et al.[35] observed higher 30-day mortality in association with low lymphocyte and albumin levels at admission in their prospective cohort study. Another study conducted in our country in 2019 showed that patients who died within the first year following hip fracture repair exhibited a higher baseline neutrophil-to-lymphocyte ratio than survivors[36]. A subsequent meta-analysis revealed a significant relationship between the neutrophil-to-lymphocyte ratio and mortality at 1 year after hip fracture repair, but not at 30 days[37]. However, in the present study, we determined that for each one-unit decrease in admission lymphocyte count, the odds of early mortality increased by 0.2%. Further research is needed

to determine the predictive value of lymphocyte count for short-term mortality when evaluating early preoperative risk factors in hip fracture patients. Moreover, Seyedi et al.[38] demonstrated that patients with high BUN levels (>20 mg/dL) and patients with high creatinine (>1.3 mg/dL) levels had a 3-fold and 2.5-fold higher risk of mortality in the first 3 months, respectively.

Study Limitations

Certain limitations of this study must be acknowledged. First, the analysis was retrospective and included a small number of patients from a single center. Other factors that may be associated with mortality were not analyzed, such as the stage of dementia, malnutrition, and medication use. However, an important strength of this research is its contribution to the literature on predictors of early mortality following geriatric hip fractures in the Turkish population.

Conclusion

Older adults face a substantial risk of mortality following a hip fracture. Operative intervention significantly improved survival, whereas admission lymphopenia, comorbid dementia, and pertrochanteric fracture locations were associated with higher early mortality rates. Therefore, we recommend close monitoring in patients presenting with these specific vulnerabilities.

Ethics

Ethics Committee Approval: Prior to the commencement of the study, formal approval was obtained from the Clinical Research Ethics Committee of Atatürk University Faculty of Medicine (approval no: 06, date: 21.02.2024).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

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

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

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

References

1. Ensrud KE. Epidemiology of fracture risk with advancing age. J Gerontol A Biol Sci Med Sci. 2013;68:1236-42.
2. Krickl J, Ittermann T, Thannheimer A, et al. The rising problem of hip fractures in geriatric patients-analysis of surgical influences on the outcome. J Pers Med. 2023;13:1271.
3. Cooper C, Campion G, Melton LJ 3rd. Hip fractures in the elderly: a world-wide projection. Osteoporos Int. 1992;2:285-9.
4. Cheng SY, Levy AR, Lefaivre KA, Guy P, Kuramoto L, Sobolev B. Geographic trends in incidence of hip fractures: a comprehensive literature review. Osteoporos Int. 2011;22:2575-86.
5. Gullberg B, Johnell O, Kanis JA. World-wide projections for hip fracture. Osteoporos Int. 1997;7:407-13.
6. Aharonoff GB, Koval KJ, Skovron ML, Zuckerman JD. Hip fractures in the elderly: predictors of one year mortality. J Orthop Trauma. 1997;11:162-5.
7. Penrod JD, Litke A, Hawkes WG, et al. The association of race, gender, and comorbidity with mortality and function after hip fracture. J Gerontol A Biol Sci Med Sci. 2008;63:867-72.
8. Hasegawa Y, Suzuki S, Wingstrand H. Risk of mortality following hip fracture in Japan. J Orthop Sci. 2007;12:113-7.
9. Ireland AW, Kelly PJ, Cumming RG. Risk factor profiles for early and delayed mortality after hip fracture: analyses of linked Australian department of veterans' affairs databases. Injury. 2015;46:1028-35.
10. Härstedt M, Rogmark C, Sutton R, Melander O, Fedorowski A. Impact of comorbidity on 6-month hospital readmission and mortality after hip fracture surgery. Injury. 2015;46:713-8.
11. Uzoigwe CE, Burnand HG, Cheesman CL, Aghedo DO, Faizi M, Middleton RG. Early and ultra-early surgery in hip fracture patients improves survival. Injury. 2013;44:726-9.
12. Colais P, Di Martino M, Fusco D, Perucci CA, Davoli M. The effect of early surgery after hip fracture on 1-year mortality. BMC Geriatr. 2015;15:141.
13. Aqil A, Hossain F, Sheikh H, Aderinto J, Whitwell G, Kapoor H. Achieving hip fracture surgery within 36 hours: an investigation of risk factors to surgical delay and recommendations for practice. J Orthop Traumatol. 2016;17:207-13.
14. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis. 1987;40:373-83.
15. Hawkes WG, Wehren L, Orwig D, Hebel JR, Magaziner J. Gender differences in functioning after hip fracture. J Gerontol A Biol Sci Med Sci. 2006;61:495-9.
16. Cree M, Soskolne CL, Belseck E, et al. Mortality and institutionalization following hip fracture. J Am Geriatr Soc. 2000;48:283-8.
17. Ensrud KE, Ewing SK, Taylor BC, et al. Frailty and risk of falls, fracture, and mortality in older women: the study of osteoporotic fractures. J Gerontol A Biol Sci Med Sci. 2007;62 744-51.
18. Hamerman D. Toward an understanding of frailty. Ann Intern Med. 1999;130:945-50.
19. Ferrucci L, Guralnik JM, Studenski S, et al. Designing randomized, controlled trials aimed at preventing or delaying functional decline and disability in frail, older persons: a consensus report. J Am Geriatr Soc. 2004;52:625-34.
20. Miller RR, Cappola AR, Shardell MD, et al. Persistent changes in interleukin-6 and lower extremity function following hip fracture. J Gerontol A Biol Sci Med Sci. 2006;61:1053-8.

21. Wehren LE, Hawkes WG, Orwig DL, Hebel JR, Zimmerman SI, Magaziner J. Gender differences in mortality after hip fracture: the role of infection. J Bone Miner Res. 2003;18:2231-7.
22. Abrahamsen B, van Staa T, Ariely R, Olson M, Cooper C. Excess mortality following hip fracture: a systematic epidemiological review. Osteoporos Int. 2009;20:1633-50.
23. Ruggiero C, Bonamassa L, Pelini L, et al. Early post-surgical cognitive dysfunction is a risk factor for mortality among hip fracture hospitalized older persons. Osteoporos Int. 2017;28:667-75.
24. Ha YC, Cha Y, Yoo JI, Lee J, Lee YK, Koo KH. Effect of dementia on postoperative mortality in elderly patients with hip fracture. J Korean Med Sci. 2021;36:e238.
25. Lai YC, Tang PL, Kuo TJ, Hsu CJ. Different impacts of dementia on two-year mortality after osteosynthesis and hemiarthroplasty in treating geriatric hip fractures. Arch Gerontol Geriatr. 2018;79:116-22.
26. Bai J, Zhang P, Liang X, Wu Z, Wang J, Liang Y. Association between dementia and mortality in the elderly patients undergoing hip fracture surgery: a meta-analysis. J Orthop Surg Res. 2018;13:298.
27. Saul D, Riekenberg J, Ammon JC, Hoffmann DB, Sehmisch S. Hip fractures: therapy, timing, and complication spectrum. Orthop Surg. 2019;11:994-1002.
28. Barceló M, Torres OH, Mascaró J, Casademont J. Hip fracture and mortality: study of specific causes of death and risk factors. Arch Osteoporos. 2021;16:15.
29. Parker MJ, Gurusamy K. Internal fixation versus arthroplasty for intracapsular proximal femoral fractures in adults. Cochrane Database Syst Rev. 2006;2006:CD001708.
30. Pereira SR, Puts MT, Portela MC, Sayeg MA. The impact of prefracture and hip fracture characteristics on mortality in older persons in Brazil. Clin Orthop Relat Res. 2010;468:1869-83.
31. Johnston AT, Barnsdale L, Smith R, Duncan K, Hutchison JD. Change in long-term mortality associated with fractures of the hip: evidence from the scottish hip fracture audit. J Bone Joint Surg Br. 2010;92:989-93.
32. Ricci G, Longaray MP, Gonçalves RZ, Neto Ada S, Manente M, Barbosa LB. Evaluation of the mortality rate one year after hip fracture and factors relating to diminished survival among elderly people. Rev Bras Ortop. 2015;47:304-9.
33. Moja L, Piatti A, Pecoraro V, et al. Timing matters in hip fracture surgery: patients operated within 48 hours have better outcomes. A meta-analysis and meta-regression of over 190,000 patients. PLoS ONE. 2012;7:e46175.
34. Lizaur-Utrilla A, Gonzalez-Navarro B, Vizcaya-Moreno MF, Miralles Muñoz FA, Gonzalez-Parreño S, Lopez-Prats FA. Reasons for delaying surgery following hip fractures and its impact on one year mortality. Int Orthop. 2019;43:441-8.
35. Blanco JF, da Casa C, Pablos-Hernández C, González-Ramírez A, Julián-Enríquez JM, Díaz-Álvarez A. 30-day mortality after hip fracture surgery: influence of postoperative factors. PLoS One. 2021;16:e0246963.
36. Temiz A, Ersözlü S. Admission neutrophil-to-lymphocyte ratio and postoperative mortality in elderly patients with hip fracture. Ulus Travma Acil Cerrahi Derg. 2019;25:71-4.
37. Chen YH, Chou CH, Su HH, et al. Correlation between neutrophil-to-lymphocyte ratio and postoperative mortality in elderly patients with hip fracture: a meta-analysis. J Orthop Surg Res. 2021;16:681.
38. Seyedi HR, Mahdian M, Khosravi G, et al. Prediction of mortality in hip fracture patients: role of routine blood tests. Arch Bone Jt Surg. 2015;3:51-5.

DOI: 10.4274/anatoljmed.2025.93898
Anatol J Gen Med Res 2026;36(2):118-23

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

The Role of Placental Elastography in Predicting PAPP-A Levels: An Observational Cross-sectional Study

PAPP-A Düzeylerinin Tahmininde Plasental Elastografinin Rolü: Gözlemsel Kesitsel Bir Çalışma

Halime Şen Selim[1], Mustafa Şengül[2], Yeşim Akgün[1], Çağlasu Sancaktar[1]

[1]İzmir Atatürk Training and Research Hospital, Clinic of Obstetrics and Gynecology, İzmir, Türkiye
[2]İzmir Katip Çelebi University Faculty of Medicine, Department of Obstetrics and Gynecology, İzmir, Türkiye

Cite as: Şen Selim H, Şengül M, Akgün Y, Sancaktar Ç. The role of placental elastography in predicting PAPP-A levels: an observational cross-sectional study. Anatol J Gen Med Res. 2026;36(2):118-23

Abstract

Objective: Pregnancy-associated plasma protein A (PAPP-A) is synthesized by placental trophoblastic tissues. Sonoelastography is a non-invasive ultrasonographic technique for assessing tissue stiffness. Studies have recently highlighted the clinical significance of evaluating placental elasticity to determine its function in obstetrics. Thus, the present study examines whether placental elastography can detect pathological changes in placental tissue linked to serum PAPP-A levels.

Methods: In an observational cross-sectional study, we evaluated 200 pregnant women at 10-14 weeks' gestation and performed ultrasound elastography examinations of the placenta. Serum PAPP-A levels (mIU/L) and nuchal translucency thickness (mm) were recorded. Correlation analysis was performed to assess the relationship between placental elastography-placental strain ratio (PSR) values and PAPP-A levels.

Results: The patients' median age was 29 years [interquartile range (IQR), 17-42]. Median values were: the crown-rump length, 59.1 (IQR 45-84) mm; nuchal thickness, 1.0 (IQR 0.47-2.0) mm; and PAPP-A, 0.96 (IQR) mIU/L. No significant correlation was found between PSR value and PAPP-A (rho=-0.118, p=0.104).

Conclusion: In the correlation analysis, we could not find a statistically significant association between placental elastography PSR values and PAPP-A. According to the first-trimester combined test results, only 3 of our patients were in the high-risk group. We used strain elastography software. Nevertheless, shear-wave elastography may yield more sensitive results. These could be limitations of our study and possible reasons why we did not detect a significant relationship.

Keywords: Pregnancy-associated plasma protein A, placental strain elastography, ultrasound

Öz

Amaç: Gebelikle ilişkili plazma proteini A (PAPP-A), plasental trofoblastik dokular tarafından sentezlenir. Sonoelastografi, doku sertliğini non-invaziv ultrasonografi kullanarak teşhis etmek için kullanılan bir tekniktir. Son zamanlarda yapılan çalışmalarda plasentanın elastikiyetinin değerlendirilmesinin doğumdaki işlevini belirlemede klinik önemi vurgulanmıştır. Bu nedenle, bu çalışmada plasental elastografinin serum PAPP-A düzeyi ile ilişkili plasenta dokusundaki patolojik değişiklikleri tespit edip edemeyeceği araştırılmıştır.

Address for Correspondence/Yazışma Adresi: Yeşim Akgün, MD, İzmir Atatürk Training and Research Hospital, Clinic of Obstetrics and Gynecology, İzmir, Türkiye
E-mail: yesimakgun35@gmail.com
ORCID ID: orcid.org/0000-0001-7486-6075

Received/Geliş tarihi: 14.10.2025
Accepted/Kabul tarihi: 27.11.2025
Epub: 02.07.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

Yöntem: Gözlemsel kesitsel bir çalışmada, 10-14. gebelik haftalarında 200 gebe kadın değerlendirildi ve plasentaya ultrasonografi-elastografi incelemesi yapıldı. Serum PAPP-A düzeyleri (mIU/L) ve ense kalınlığı (mm) kaydedildi. Plasental elastografi-plasental gerinim oranı (PSR) değerleri ile PAPP-A düzeyleri arasında korelasyon analizi yapıldı.

Bulgular: Hastaların yaş medyanı 29 [çeyrekler arası aralık (ÇAA), 17-42] yıldı. Baş-popo mesafesinin ortanca değeri 59,1 (ÇAA 45-84) mm, NT 1,0 (ÇAA 0,47-2,0) mm ve PAPP-A 0,96 (ÇAA) mIU/L olarak bulundu. PSR değeri ile PAPP-A arasında anlamlı bir korelasyon bulunmadı (rho=-0,118 p=0,104).

Sonuç: Korelasyon analizinde plasental elastografi PSR değerleri ile PAPP-A arasında istatistiksel olarak anlamlı bir korelasyon bulamadık. Birinci trimester kombine test sonuçlarına göre hastalarımızın sadece 3'ü yüksek risk grubundaydı ve strain elastografi yazılımını kullandık. Yine de, share-wave elastografisi daha hassas sonuçlar verebilir. Bunlar çalışmamızın kısıtlılıkları olabilir ve anlamlı bir ilişki bulamamamızın olası nedenleridir.

Anahtar Kelimeler: Gebelik ilişkili plazma proteini A, plasental elastografi, ultrason

Introduction

Pregnancy-associated plasma protein A (PAPP-A) is synthesized by placental trophoblastic tissue and is detected in maternal serum[1]. It is most commonly used as a component of the combined first-trimester test for Down syndrome screening. It is detected at low levels in trisomies 21, 18, and 13[2,3]. Moreover, many studies have described its relationship with undesirable complications caused by abnormal placentation, such as preterm delivery, intrauterine growth retardation (IUGR), preeclampsia (PE), miscarriage, stillbirth, and placenta accreta[4-7].

PAPP-A in maternal serum is primarily produced and released by the syncytiotrophoblast[8]. Various studies indicate that alterations in placental tissues are linked to PAPP-A levels. The findings reveal a marked decrease in the size of terminal villi in pregnancies with reduced PAPP-A levels and fetal growth restriction compared to a control group with average PAPP-A levels and uncomplicated pregnancies[9]. This is a crucial indication that should not be overlooked. Placental pathology has been associated with non-optimal first-trimester secretion of the analyte PAPP-A in pregnancies with adverse outcomes, particularly PE[10]. Also, in pregnancies affected by T21, there is a notable presence of both fetal structural abnormalities, which manifest as various physical malformations, and placental developmental defects, hich can compromise the vital functions of the placenta and impact the well-being of the fetus. Histopathological evaluations of placentas affected by T21 reveal striking characteristics: pronounced enlargement, asymmetry, and disordered arrangement of chorionic villi. These marked abnormalities underscore the profound effects of the condition on placental structure and function[11]. In addition, recent studies have revealed that abnormalities in trophoblast fusion and differentiation can significantly affect placental development. These defects lead to a higher proportion of trophoblasts displaying a double-layered structure, which can disrupt normal placental function. Additionally, an increased number of villous capillaries may contribute to altered blood flow and nutrient exchange in the placenta. This atypical vascularization may have implications for fetal growth and overall pregnancy outcomes[11].

Sonoelastography is a technique for diagnosing tissue stiffness using non-invasive ultrasonography[12]. This method is commonly employed to evaluate the biomechanical alterations in affected parenchymal tissue[12,13]. Studies have recently highlighted the clinical significance of evaluating placental elasticity to determine its function in obstetrics[14-16]. Thus, the present study examines whether placental elastography, a non-invasive functional method, can detect pathological changes in placental tissue linked to the serum PAPP-A level. Our goal was to establish the association between serum PAPP-A levels and placental elastography values.

Materials and Methods

This study was reviewed and approved by the İzmir Katip Çelebi University Clinical Research Ethics Committee (approval no: 16, date: 23.02.2023). The study was conducted in accordance with the principles outlined in the Helsinki Declaration. To ensure compliance, all patients provided signed consent forms.

Our study was conducted prospectively as an observational cross-sectional study. We did not enroll patients with pre-existing systemic diseases, such as diabetes mellitus, hypertension, cardiac conditions, or thyroid conditions.

However, we included all other pregnant women who voluntarily agreed to participate in the study.

A power analysis for binary logistic regression determined that the minimum sample size was 192 individuals. In this case, the power of the test is expected to be approximately 80.1%. We evaluated the first 200 pregnant women at 10-14 weeks of gestation who applied to our hospital and consented to participate in the study between February 23 to April 30, 2023, consecutively.

We employed the Esaote MyLabSeven real-time tissue elastography software and an 8-1 MHz multi-frequency AC2541 probe for placental assessment during a routine first-trimester fetal ultrasound evaluation. The medical practitioner employed the ultrasound probe to apply and release pressure on the tissue during the placental elastography examination, thereby compressing and decompressing it. The compression process ensures that the transducer optimizes the machine's color coding.

Tissue elasticity was represented on the screen by a color scale (red, yellow, green, and blue). Red indicates the lowest elasticity, whereas blue indicates the highest. A total of 4 regions of interest (ROIs), each with an area of 31.89 mm^2, were determined. One area (A, circle Z1) was located in the myometrium, and 3 areas (B, circles Z2, Z3, and Z4) were located in the placenta. When choosing ROIs, we selected areas with homogeneous distributions of blue, green, and red colors. The device automatically calculated the ratios of elasticity among these four areas (Elx1, Elx2, Elx3, and Elx4) and determined the SR values (Elx2/Elx1=SR1, Elx3/Elx1=SR2, and Elx4/Elx1=SR3) (Figure 1). We have defined the placental strain ratio (PSR) as the average strain ratio (SR1+SR2+SR3)/3.

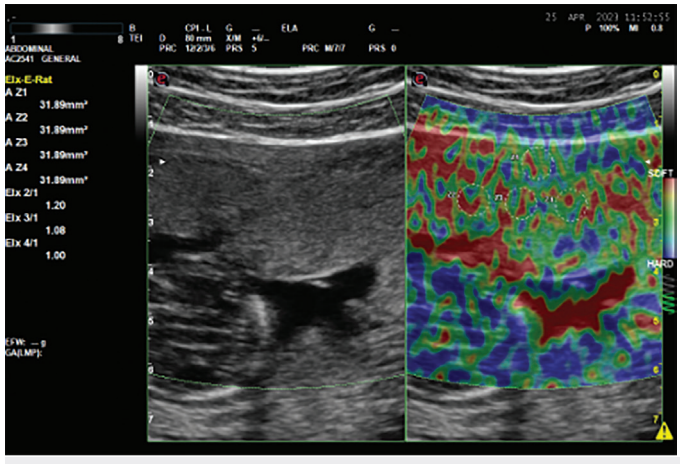


Figure 1. Ultrasound image of placental sonoelastography

Measurements were taken three separate times by two different blinded researchers and averaged to avoid potential bias.

The first-trimester combined test results, serum PAPP-A levels (mIU/L), and nuchal translucency (NT) thickness (mm) were extracted from our hospital's patient records.

Statistical Analysis

We used SPSS software (version 23.0) (IBM Corp., Armonk, N.Y., USA) also known as the Statistical Package for the Social Sciences. The Kolmogorov-Smirnov test was used to assess the normality of the distributions. Correlation analysis between placental elastography-placental ratio (PSR) values and PAPP-A levels was performed to evaluate the extent to which PSR reflected PAPP-A values. We examined the correlation between placental elastography values and PAPP-A values using Spearman correlation analysis and determined the correlation coefficient (rho). If the p-value was less than 0.05, it was considered statistically significant.

Results

The patients' ages ranged from 17 to 42 years, with a median age of 29 years. The crown-rump length (CRL) at the time of the first-trimester combined testing was 59.1 mm (45-84 mm); NT was 1.0 mm (0.47-2.0 mm); and PAPP-A had a median of 0.96 mIU/L (Table 1). The first-trimester combined test results for 3 patients indicated high risk (trisomy 21 risks were 1:42, 1/131, and 1/135, respectively). The risk of trisomy 21 for 12 pregnant women was 1/250-1/1000, while the remaining pregnant women were in the low-risk range (1/1216-1/99000).

Amniocentesis was performed on the pregnant woman with a risk of 1/42, and the result indicated a normal karyotype.

Table 1. Demographic variables

Variables	Data
Demographic variables (n=200)	
Age (year), median (IQR)	29 (8)
Gravida, median (IQR)	2 (2)
Parity, median (IQR)	1 (1)
CRL (mm), median (IQR)	59.1 (12.4)
NT (mm), median (IQR)	1.0 (0.2)
PAPP-A (mIU/L), median (IQR)	0.96 (0.77)
Placental strain ratio (PSR), median (IQR)	0.79 (0.79)

CRL: Crown-rump length, NT: Nuchal thickness, PAPP-A: Pregnancy-associated plasma protein A, IQR: Interquartile range

The other two refused to undergo an invasive procedure, but no anomalies were detected on second-trimester ultrasound examinations in any of our patients.

PAPP-A levels ranged from 0.17 to 6.9 mIU/L, with a median of 0.96 mIU/L. The values of PSR varied from 0.6 to 1.5, with a median of 0.79 (Figure 2).

Correlation analysis revealed no notable correlation between the PSR value measured by placental elastography and PAPP-A. (rho=-0.118 p=0.104) (Table 2).

Discussion

Pregnancy complications can cause fetal and maternal morbidities, so finding the causes of these complications and detecting them early have always been focuses of attention among researchers. Many parameters have been scrutinized in this regard. One of them is the maternal serum PAPP-A level. Many studies have evaluated the relationship between PAPP-A levels and pregnancy complications, such as preterm delivery, IUGR, PE, miscarriage, and stillbirth[4-7].

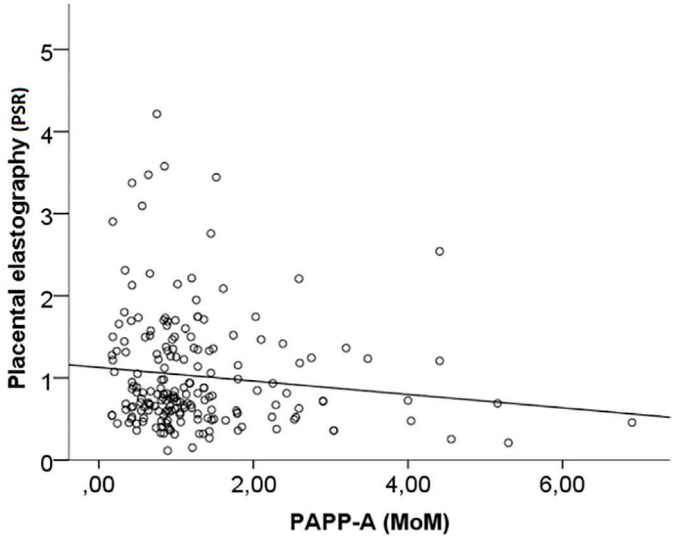


Figure 2. Correlation of PAPP-A and PSR

PAPP-A: Pregnancy-associated plasma protein A, PSR: Plasenta strain ratio, MoM: Multiples of the median

Table 2. Correlation analysis of PAPP-A with PSR value

Variables	rho	p-value
PAPP-A	-0.118	0.104

rho: Spearman's rank correlation coefficient, PAPP-A: Pregnancy-associated plasma protein A, PSR: Plasenta strain ratio

Rojas-Rodriguez et al.[16] stated that PAPP-A is a metalloprotease secreted by the placenta and modulates the bioavailability of insulin-like growth factor (IGF) through proteolysis of IGF-binding proteins (IGFBPs) 2, 4, and 5. The researchers discovered that mice lacking PAPP-A exhibited conditions similar to those observed in humans with gestational diabetes mellitus, such as adipose tissue restructuring, pregnancy-induced insulin resistance, and fatty liver disease[17].

Boutin et al.[17] conducted a study to assess how well PAPP-A can predict placental outcomes in low-risk nulliparous pregnancies. PAPP-A is moderately associated with PE [with an area under the curve (AUC) of 0.57 and a 95% confidence interval (CI) of 0.53-0.61] and with small-for-gestational-age infants (with an AUC of 0.62 and a 95% CI: of 0.56-0.69). However, fetal death was not associated with PAPP-A (with an AUC of 0.43 and a 95% CI: of 0.23-0.63)[18]. Odibo et al.[9] found significant differences in first-trimester PAPP-A levels of pregnant women who experienced PE, growth restriction, or preterm delivery later in their pregnancy. Kapustin et al.[18] evaluated PAPP-A and fb-hCG values in the first-trimester of pregnancies complicated by diabetes to assess their ability to predict potential obstetrical complications such as macrosomia, PE, IUGR, and preterm birth. Compared with the control group, PAPP-A and fb-hCG MoM levels were significantly lower in the 5-10% range and notably higher in the 95% range[19].

PAPP-A has been investigated in various components of the fetoplacental unit because of its association with adverse obstetric outcomes. For instance, *in vitro* study, Kavoussi et al.[19] assessed the presence of PAPP-A in blastocoel-fluid-conditioned media obtained from IVF/ICSI patients. They detected PAPP-A mRNA in 56.3% of samples[20].

Because PAPP-A is a placenta-derived hormone, researchers have also focused on placental pathology. In another study, Odibo et al.[10] discovered noteworthy variations in PAPP-A, ADAM12, and placental growth factor levels in cases of PE, premature delivery, and small-for-gestational-age infants with specific placental histological findings. Their study provided evidence linking placental pathology and inadequate secretion of analytes during the first-trimester to unfavorable outcomes, particularly PE[10]. However, pathological examination of the placenta cannot be used as a predictor during pregnancy, since it is typically performed after delivery. Therefore, the researchers aimed to predict the relationship between different ultrasound

measurements, PAPP-A levels, and indirect pregnancy complications.

Various ultrasound measurements have been examined to predict PAPP-A levels and, thus, associated complications. Wortelboer et al.[20] measured uterine artery impedance by transvaginal Doppler ultrasound and evaluated its relationship with the PAPP-A level. They concluded that the increase in these hormone levels was negatively correlated with uterine artery impedance[21]. Another study investigated the correlation between placental volume and PAPP-A levels and reported that PAPP-A levels were closely associated with placental size up to 8 weeks. This relationship disappeared after the luteoplacental shift at 8 weeks[22].

Papastefanou et al.[22] evaluated the correlation between PAPP-A and first-trimester fetal ultrasound parameters, including dCRL, fetal heart rate (dFHR), and dNT. They found that PAPP-A had a direct correlation with dNT and an inverse correlation with dCRL and dFHR. The statistical analysis revealed a significant difference between the observed and adjusted MoM, with a p-value of less than 0.001[23].

Our objective was to establish a conclusive link between PAPP-A levels and placental elastography, a non-invasive ultrasound method that can indirectly indicate pathological changes in the placenta by assessing tissue stiffness. To our knowledge, this is the first study using such an approach. Generally, studies on ultrasound elastography have focused on the pathology of tissues, such as the thyroid and breast[24,25]. However, recent ultrasound elastography studies have also focused on pregnant women to investigate pregnancy complications such as pregnancy-induced hypertension, PE, and IUGR[26-29]. According to a literature review by Edwards et al.[25], studies have shown a correlation between PSR values and pregnancy complications. Ultrasound elastography can provide a quantitative evaluation of the placenta's biomechanical characteristics when its function is affected[25].

The association of first-trimester placental biomarkers with adverse pregnancy outcomes continues to attract researchers' attention. This recent study once again highlighted that when combinations of biomarkers were assessed, PAPP-A and AFP were independently associated with small-for-gestational-age, and that PAPP-A alone was associated with pre-eclampsia and preterm birth[29].

Study Limitations

Our correlation analysis did not reveal a meaningful association between placental elastography PSR values and PAPP-A. According to the first-trimester combined test results, only 3 of our patients were in the high-risk group. This could be a limitation of our study and a possible reason why we did not find a significant relationship. We also performed ultrasonography in the first-trimester, but PSR values may differ in the third trimester. This timing limitation must be considered in our study. Lastly, we used strain elastography software, but shear-wave elastography may yield more sensitive measurements.

PAPP-A has been the subject of studies since the 1970s, when it was first defined. Its new functions are still being determined, and research on it has continued for 50 years[1]. Future studies are likely to reveal findings elucidating the mechanisms of PAPP-A.

Conclusion

Correlation analysis revealed no statistically significant association between placental elastography PSR values and PAPP-A levels. A potential limitation of our study is the small number of high-risk cases identified based on the first-trimester combined test, as only three patients fell into this category. Furthermore, the use of strain elastography software may have contributed to the lack of significant findings, as shear-wave elastography could yield more precise and sensitive results. These factors may collectively explain the absence of a significant correlation.

In prospective studies involving larger patient cohorts, second- and third-trimester elastography examinations could be included, which might lead to more definitive conclusions regarding the PAPP-A level and the prediction of pregnancy complications. We believe that our study may prompt further research.

Ethics

Ethics Committee Approval: This study was reviewed and approved by the İzmir Katip Çelebi University Clinical Research Ethics Committee (approval no: 16, date: 23.02.2023).
Informed Consent: Written informed consent was obtained from all participants.

Footnotes

Authorship Contributions

Surgical and Medical Practises: H.Ş.S., Ç.S., Concept: H.Ş.S., M.Ş., Design: Y.A., Ç.S., Data Collection or Processing: H.Ş.S., Ç.S., Analysis or Interpretation: H.Ş.S., M.Ş., Literature Search: M.Ş., Ç.S., Writing: H.Ş.S., Y.A.

Conflict of Interest: The authors declare no conflict of interest.

Financial Disclosure: This research received no external funding.

References

1. Conover CA, Oxvig C. The pregnancy-associated plasma protein-A (PAPP-A) story. Endocr Rev. 2023;44:1012-28.
2. Ziolkowska K, Tobola-Wrobel K, Dydowicz P, Zurawski S, Pietryga M, Wysocka E. The significance of maternal blood pregnancy-associated plasma protein A (PAPP-A) and free beta-subunit of human chorionic gonadotropin (β-hCG) levels for the risk assessment of fetal trisomy 18 during the first prenatal testing between 11 and 13+6 weeks of pregnancy. Ginekol Pol. 2020;91:748-4.
3. Ziolkowska K, Dydowicz P, Sobkowski M, Tobola-Wrobel K, Wysocka E, Pietryga M. The clinical usefulness of biochemical (free β-hCg, PaPP-a) and ultrasound (nuchal translucency) parameters in prenatal screening of trisomy 21 in the first trimester of pregnancy. Ginekol Pol. 2019;90:161-6.
4. Wang F, Chen S, Wang J, et al. First trimester serum PAPP-A is associated with placenta accreta: a retrospective study. Arch Gynecol Obstet. 2021;303:645-2.
5. Zafman KB, Getrajdman CS, Arnold MK, Stone JL. Are Women with a history of low PAPP-A at risk for adverse perinatal outcomes in a subsequent pregnancy? Am J Perinatol. 2019;36:647-2.
6. Prasad P, Romero R, Chaiworapongsa T, et al. Further evidence that an episode of premature labor is a pathologic state: involvement of the insulin-like growth factor system. Fetal Diagn Ther. 2023;50:236-7.
7. Honarjoo M, Kohan S, Zarean E, Tarrahi MJ. Assessment of β-human-derived chorionic gonadotrophic hormone (βhCG) and pregnancy-associated plasma protein A (PAPP-A) levels as predictive factors of preeclampsia in the first trimester among Iranian women: a cohort study. BMC Pregnancy Childbirth. 2019;19:464.
8. Livrinova V, Petrov I, Samardziski I, et al. Clinical importance of low level of PAPP-A in first trimester of pregnancy - an obstetrical dilemma in chromosomally normal fetus. Open Access Maced J Med Sci. 2019;7:1475-9.
9. Odibo AO, Zhong Y, Longtine M, et al. First-trimester serum analytes, biophysical tests and the association with pathological morphometry in the placenta of pregnancies with preeclampsia and fetal growth restriction. Placenta. 2011;32:333-8.
10. Odibo AO, Patel KR, Spitalnik A, Odibo L, Huettner P. Placental pathology, first-trimester biomarkers and adverse pregnancy outcomes. J Perinatol. 2014;34:186-91.
11. Kloth C, Kratzer W, Schmidberger J, Beer M, Clevert DA, Graeter T. Ultrasound 2020 - diagnostics & therapy: on the way to multimodal ultrasound: contrast-enhanced ultrasound (CEUS), microvascular Doppler techniques, fusion imaging, sonoelastography, interventional sonography. Rofo. 2021;193:23-32.
12. Bodard S, Lassalle L, Larousserie F, et al. Performance of sonoelastography for predicting malignancy in soft tissue. BMC Cancer. 2022;22:1305.
13. Eroğlu H, Tolunay HE, Tonyalı NV, Orgul G, Şahin D, Yücel A. Comparison of placental elasticity in normal and intrauterine growth retardation pregnancies by ex vivo strain elastography. Arch Gynecol Obstet. 2020;302:109-5.
14. Fujita Y, Nakanishi TO, Sugitani M, Kato K. Placental elasticity as a new non-invasive predictive marker of pre-eclampsia. Ultrasound Med Biol. 2019;45:93-7.
15. Saw SN, Low JYR, Mattar CNZ, Biswas A, Chen L, Yap CH. Motorizing and optimizing ultrasound strain elastography for detection of intrauterine growth restriction pregnancies. Ultrasound Med Biol. 2018;44:532-3.
16. Rojas-Rodriguez R, Ziegler R, DeSouza T, et al. PAPPA-mediated adipose tissue remodeling mitigates insulin resistance and protects against gestational diabetes in mice and humans. Sci Transl Med. 2020;12:eaay4145.
17. Boutin A, Gasse C, Demers S, Blanchet G, Giguère Y, Bujold E. Does low PAPP-A predict adverse placenta-mediated outcomes in a low-risk nulliparous population? The great obstetrical syndromes (GOS) study. J Obstet Gynaecol Can. 2018;40:663-8.
18. Kapustin RV, Kascheeva TK, Alekseenkova EN, Shelaeva EV. Are the first-trimester levels of PAPP-A and fb-hCG predictors for obstetrical complications in diabetic pregnancy? J Matern Fetal Neonatal Med. 2022;35:1113-9.
19. Kavoussi SK, Chen SH, Wininger JD, et al. The expression of pregnancy-associated plasma protein-A (PAPP-A) in human blastocoel fluid-conditioned media: a proof of concept study. J Assist Reprod Genet. 2022;39:389-4.
20. Wortelboer EJ, Koster MP, Kuc S, et al. Longitudinal trends in fetoplacental biochemical markers, uterine artery pulsatility index and maternal blood pressure during the first trimester of pregnancy. Ultrasound Obstet Gynecol. 2011;38:383-8.
21. Sahraravand M, Järvelä IY, Laitinen P, Tekay AH, Ryynänen M. The secretion of PAPP-A, ADAM12, and PP13 correlates with the size of the placenta for the first month of pregnancy. Placenta. 2011;32:999-1003.
22. Papastefanou I, Souka A, Eleuftheriades M, Pilalis A, Kassanos D. First trimester fetal ultrasound parameters associated with PAPP-A and fβ-hCG. J Matern Fetal Neonatal Med. 2012;25:1029-33.
23. Alqahtani SM, Alanesi SF, Mahmood WS, et al. Clinical and ultrasonographic features in cancer risk stratification of indeterminate thyroid nodules. Saudi Med J. 2022;43:473-8.
24. Alsharif WM. The utilization of artificial intelligence applications to improve breast cancer detection and prognosis. Saudi Med J. 2023;44:119-27.
25. Edwards C, Cavanagh E, Kumar S, Clifton V, Fontanarosa D. The use of elastography in placental research - a literature review. Placenta. 2020;99:78-88.
26. Meena R, Malik A, Jain S, Batra A. Placental elastography in second trimester preeclampsia prediction: a prospective study. Ultrasound. 2022;30:228-5.
27. Hu J, Lv Z, Dong Y, Liu W. Review of shear wave elastography in placental function evaluations. J Matern Fetal Neonatal Med. 2023;36:2203792.
28. Deeba F, Hu R, Lessoway V, et al. SWAVE 2.0 imaging of placental elasticity and viscosity: potential biomarkers for placenta-mediated disease detection. Ultrasound Med Biol. 2022;48:2486-501.
29. Sovio U, Gaccioli F, Cook E, Charnock-Jones DS, Smith GCS. Association between adverse pregnancy outcome and placental biomarkers in the first trimester: a prospective cohort study. BJOG. 2024;131:823-31.

DOI: 10.4274/anatoljmed.2026.85520
Anatol J Gen Med Res 2026;36(2):124-31

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Surgical Management of Giant Intracranial Meningiomas: Key Considerations and Postoperative Results

Dev İntrakraniyal Menenjiyomların Cerrahi Tedavisi: Temel Hususlar ve Ameliyat Sonrası Sonuçlar

Yücel Doğruel[1], Muhammet Enes Gürses[2], Elif Gökalp[2], Oğuz Kağan Demirtaş[3], Aleyna Gökçen Şahin[4], Bülent Demirgil[4], Erhan Emel[4], Abuzer Güngör[5]

[1]University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Neurosurgery, İzmir, Türkiye
[2]Tatvan State Hospital, Clinic of Neurological Surgery, Bitlis, Türkiye
[3]University of Health Sciences Türkiye, Ankara Etlik City Hospital, Department of Neurosurgery, Ankara, Türkiye
[4]University of Health Sciences Türkiye, Bakırköy Prof. Dr. Mazhar Osman Mental Health and Neurological Diseases Training and Research Hospital, Department of Neurosurgery, İstanbul, Türkiye
[5]İstinye University Faculty of Medicine, Department of Neurosurgery, İstanbul, Türkiye

Cite as: Doğruel Y, Gürses ME, Gökalp E, et al. Surgical management of giant intracranial meningiomas: key considerations and postoperative results. Anatol J Gen Med Res. 2026;36(2):124-31

Abstract

Objective: Giant intracranial meningiomas (GIM) present significant surgical challenges due to their size and involvement with critical neurovascular structures. This study aims to evaluate key surgical considerations and postoperative outcomes for patients undergoing resection of GIM and to highlight factors influencing recurrence-free survival and patient prognosis.

Methods: A retrospective review was conducted of 8 patients with GIM (>5 cm in diameter) who underwent surgical resection between 2021 and 2023. Demographic data, tumor characteristics, surgical techniques, and postoperative outcomes were analyzed. Tumor size, extent of resection, and recurrence were assessed through imaging and clinical follow-up.

Results: The cohort included 3 males (37.5%) and 5 females (62.5%) with a mean age of 56.4 years. The most common tumor location was the olfactory groove (37.5%). The mean tumor volume was 107.6 cm^3, and the mean follow-up duration was 9.6 months. Gross total resection (Simpson grade 2) was achieved in 75% of cases, while subtotal resection (Simpson grade 4) was achieved in 25% of cases. Two patients with World Health Organization grade 2 pathology received postoperative radiotherapy. No major postoperative complications were observed, and the mean duration of hospitalization was 11 days.

Conclusion: Surgical resection of GIM remains a challenging but essential approach to managing these complex tumors. While gross total resection is the preferred goal, subtotal resection may be necessary to preserve neurological function. Postoperative radiotherapy may be required in higher-grade tumors. Tumor location, vascularity, and advanced intraoperative techniques significantly influence surgical outcomes and recurrence-free survival.

Keywords: Giant intracranial meningioma, recurrence-free survival, surgical outcome, meningioma

Address for Correspondence/Yazışma Adresi: Prof., Abuzer Güngör, İstinye University Faculty of Medicine, Department of Neurosurgery, İstanbul, Türkiye
E-mail: abuzergungor@gmail.com
ORCID ID: orcid.org/0000-0002-2792-7610

Received/Geliş tarihi: 13.11.2025
Accepted/Kabul tarihi: 19.01.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

Amaç: Dev intrakraniyal menenjiyomlar (GİM), boyutları ve kritik nörovasküler yapılarla ilişkileri nedeniyle önemli cerrahi zorluklar oluşturur. Bu çalışmanın amacı, GİM rezeksiyonu uygulanan hastalarda temel cerrahi faktörleri ve postoperatif sonuçları değerlendirmek, nüksüz sağkalım ve hasta prognozunu etkileyen faktörleri ortaya koymaktır.

Yöntem: 2021-2023 yılları arasında cerrahi rezeksiyon uygulanan, çapı 5 cm'den büyük 8 GİM hastasının geriye dönük incelemesi yapıldı. Demografik veriler, tümör özellikleri, cerrahi teknikler ve postoperatif sonuçlar analiz edildi. Tümör boyutu, rezeksiyon derecesi ve nüks, görüntüleme ve klinik takipler aracılığıyla değerlendirildi.

Bulgular: Kohortta 3 erkek (%37,5) ve 5 kadın (%62,5) hasta yer aldı; ortalama yaş 56,4 idi. En sık görülen tümör yerleşimi olfaktör oluk (%37,5) idi. Ortalama tümör hacmi 107,6 cm^3, ortalama takip süresi ise 9,6 ay olarak bulundu. Olguların %75'inde total rezeksiyon (Simpson evre 2), %25'inde subtotal rezeksiyon (Simpson evre 4) gerçekleştirildi. Dünya Sağlık Örgütü evre 2 patolojiye sahip iki hastaya postoperatif radyoterapi uygulandı. Majör postoperatif komplikasyon gözlenmedi ve ortalama hastanede yatış süresi 11 gündü.

Sonuç: GİM'lerin cerrahi rezeksiyonu, bu karmaşık tümörlerin tedavisinde zorlu ancak vazgeçilmez bir yaklaşımdır. Total rezeksiyon tercih edilen hedef olsa da, nörolojik fonksiyonun korunması amacıyla subtotal rezeksiyon gerekli olabilir. Yüksek dereceli tümörlerde postoperatif radyoterapi gerekebilir. Tümör yerleşimi, vaskülaritesi ve ileri intraoperatif tekniklerin kullanımı, cerrahi sonuçlar ve nüksüz sağkalımı önemli ölçüde etkilemektedir.

Anahtar Kelimeler: Dev intrakraniyal menenjiyom, nükssüz sağkalım, cerrahi sonuç, menenjiyom

Introduction

Meningiomas are the most frequently encountered benign tumors of the central nervous system and arise from arachnoid cap cells. They characteristically develop as extra-axial, slow-progressing lesions[1,2]. Among them, giant intracranial meningiomas (GIMs) constitute a clinically significant subgroup distinguished by their extensive size and tendency to involve adjacent neurovascular structures, which considerably increases surgical complexity[3,4]. Because of their indolent growth pattern, meningiomas often remain asymptomatic for prolonged periods. Clinical manifestations generally emerge as a consequence of mass effect, tumor location, or elevated intracranial pressure, and radiological identification frequently occurs only after substantial enlargement has taken place[5]. Size-based classification systems describe meningiomas larger than 3 cm as large, while lesions exceeding 5 cm at diagnosis are categorized as giant[6,7]. The management of giant meningiomas presents particular challenges. Their considerable volume contributes to raised intracranial pressure, and their intimate relationship with essential anatomical structures complicates operative planning and execution. Consequently, these tumors demand meticulous surgical strategy and advanced technical expertise[4,7]. Various treatment protocols are employed for meningiomas, ranging from simple observation and partial resection, through radiosurgery (as a primary or adjuvant therapy), to aggressive surgical resection[8-10]. Therapeutic approaches to meningiomas range from active surveillance to subtotal resection, stereotactic radiosurgery-either as a standalone modality or adjunctive therapy-and extensive microsurgical excision[6]. In cases involving large or giant lesions, operative treatment is often indicated due to significant mass effect and neurovascular compromise. Importantly, the objectives of surgical intervention extend beyond tumor removal, emphasizing not only overall survival but also preservation of postoperative functional status and quality of life[11].

Critical factors influencing the choice of surgical approach include the tumor's complex location, growth trajectory, involvement of adjacent structures, patient age, and the surgeon's experience[12]. Surgery for GIM poses unique challenges due to the tumor's large volume, increased vascularity, involvement of vital neurovascular structures, and severe cerebral edema, all of which complicate visualization and resection. Gross total resection is the primary goal in the surgical management of meningiomas. However, achieving this is not always feasible for giant meningiomas due to their size and frequent adherence to, or encasement of, critical neurovascular structures. These factors significantly increase the risk of postoperative complications, making the surgical approach more complex and requiring careful consideration of subtotal resection in some cases to preserve neurological function and patients' quality of life.

Materials and Methods

Permission for the study was obtained from the University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training

and Research Hospital Clinical Research Ethics Committee (approval no: 2025-01-35, date: 24.10.2025).

A retrospective analysis was conducted of patients who underwent surgical treatment for intracranial tumors between 2021 and 2023 and who received a histopathological diagnosis of meningioma. Within this population, only cases with a maximum tumor diameter greater than 5 cm were included, consistent with the definition of GIM.

Demographic characteristics, clinical presentation, radiological findings, intraoperative use of neuronavigation, applied microsurgical techniques, and postoperative outcomes were systematically extracted from medical records, operative reports, imaging databases, and follow-up documentation.

Histopathological evaluation was performed in accordance with the World Health Organization classification criteria[13]. Tumor volume was estimated using the formula a×b×c/2 based on radiological measurements.

Early postoperative neuroimaging was obtained within 48 hours to assess the extent of tumor resection. During follow-up, magnetic resonance imaging was scheduled at six-month intervals throughout the first postoperative year and annually thereafter.

Recurrence was defined radiologically as either the reappearance of tumor tissue at the original site after gross total resection or a measurable enlargement of residual tumor tissue, regardless of the presence or absence of clinical symptoms.

Statistical Analysis

Statistical analyses were conducted using the Statistical Package for the Social Sciences (SPSS), version 22.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were calculated.

Results

The demographic and clinical data of 8 patients who underwent surgical intervention between 2021 and 2023 were analyzed. The demographic, clinical, and histological characteristics of all patients with giant meningiomas are summarized in Table 1. The mean follow-up duration for the patients is 9.6 months (9.6±6.8 months). The cohort included 3 males (37.5%) and 5 females (62.5%), with a mean age of 56.4 years (56.4±10.4). Among the 8

Table 1. Demographic data of GIM patients along with clinical features

Patient	Age (years)/ gender	Location	Size of tumor (mm)	Symptom/sign	Operation duration (hour)	Intraoperative blood lose (cc)	Simpson grade	Pathology grade	Length of hospital stay (day)	Postoperative radiotherapy	Follow-up (month)
1	64/F	Tuberculum sella	81x64x70	Dizziness	24	1250	2	1	6	-	18
2	61/F	Olfactory groove	60x66x45	Anosmia, syncope, visual deficit	23	1200	2	1	21	-	18
3	32/M	Convexity	93x65x58	Seizure	10	1500	4	2	3	+	9
4	50/F	Sphenoid wing	67x51x60	Facial paralysis, headache	14	800	2	1	5	-	12
5	61/F	Falx	62x55x50	Headache	14	1000	4	2	4	+	3
6	55/M	Tentorial	40x40x60	Headache	10	1250	2	1	5	-	1
7	61/M	Olfactory groove	62x50x58	Anosmia, dizziness, gait disturbance, headache	16	1100	2	1	14	-	1
8	67/F	Olfactory groove	60x50x60	Dizziness, gait disturbance, headache, incontinence	13	800	2	1	30	-	15

F: Female, M: Male, mm: Millimeter, GIM: Giant intracranial meningiomas

patients, the most common tumor location was the olfactory groove (3 cases; 37.5%), followed by tuberculum sella, convexity, sphenoid wing, falx, and tentorial locations, each representing 1 case (12.5%). The mean operation duration was 15.5 hours (15.5±5.2 hours). The mean bleeding amount was 1112.5 cc (1112.5±235.7 cc). Simpson grade 2 was observed in 75% of cases, and Simpson grade 4 was observed in 25% of cases. Pathology grade 1 was seen in 75% of cases, and pathology grade 2 was seen in 25% of cases. The mean hospitalization duration was 11 days (11.0±9.2 days). The mean tumor size is 107.6 cm^3 (107.6±43.4 cm^3). Two patients with grade 2 pathology received postoperative radiotherapy. No major postoperative complications were observed. The pre and post-resection status of cases in the study cohort are illustrated (Figure 1-8).

Discussion

This study reports the clinical results of eight patients treated surgically for GIMs. Particular emphasis was placed on the operative difficulties associated with these uncommon lesions, underscoring the necessity of comprehensive anatomical knowledge and the application of refined microsurgical strategies to optimize surgical outcomes. By doing so, we aimed to emphasize the complexity of these cases and the critical factors involved in achieving safe and effective resections.

GIM typically shows a female predominance and occurs around the age of 40[3,14]. In our series, giant meningiomas were more common in female patients (62.5%). Clinical symptoms vary according to tumor location, but headaches are the most common symptom in our series (62.5%). Chan and Thompson[15] demonstrated that parasagittal or falx meningiomas are the most common tumor localization. In our series, olfactory groove meningiomas were more frequently observed; however, the small number of cases is insufficient to support generalizations. Long operation durations and blood loss play a critical role in meningioma surgeries[16]. For tumors close to vascular structures, operative duration is longer. The outcome may vary depending on the surgeon's experience. Our series includes surgeries performed by a single surgeon. In our series, the mean intraoperative blood loss is 1112.5 cc (1112.5±224.65 cc). The mean operation duration is 15.5 hours (15.5±5 hours).

Irrespective of anatomical location, the surgical management of GIMs is technically demanding. Their considerable size, marked vascularity, associated peritumoral edema, and frequent encasement or displacement of critical neurovascular structures significantly increase operative complexity (Figure 9). Consequently, the choice of surgical strategy must be tailored to the tumor's anatomical position and vascular characteristics.

In this context, Attia et al.[17] described a skull base technique for the treatment of giant anterior clinoidal meningiomas that incorporated extradural optic canal unroofing, anterior clinoidectomy, and early decompression of the optic nerve. This method improved surgical exposure, enhanced tumor removal, and minimized the likelihood of injury to adjacent neurovascular structures. Surgical challenges in GIMs include safely identifying and preserving critical neurovascular structures, particularly in medial sphenoid wing and clinoidal meningiomas. Early optic nerve decompression and extradural clinoidectomy are key techniques in these cases, especially when visual impairment is present. For olfactory groove meningiomas, both the pterional and bifrontal approaches are commonly used. Visual impairment associated with olfactory groove meningiomas is generally less severe than that observed in clinoidal lesions and shows a more favorable recovery following a bifrontal surgical approach. Alternatively, the pterional-transsylvian corridor provides several operative advantages, including early identification and control of critical neurovascular structures, limited cortical manipulation, reduced perioperative morbidity and mortality, and relatively low recurrence rates.

In our series, two patients underwent adjuvant radiotherapy because their tumors were histopathologically classified as grade II. The mean duration of follow-up was 9.6 months, while the average length of hospitalization was 11 days. The calculated mean tumor volume was 107.6 cm^3.

Certain tumor characteristics were associated with less favorable outcomes. Invasion of the paranasal sinuses and encasement of the anterior cerebral artery were associated with higher rates of postoperative complications and of recurrence. Extensive peritumoral brain edema represents another major surgical challenge in GIM cases, particularly when accompanied by pronounced mass effect or midline shift.

To address elevated intracranial pressure and facilitate safer resection, several intraoperative strategies may be employed. Controlled cerebrospinal fluid drainage through multiple dural openings and staged internal decompression using neuronavigation guidance and a cavitron ultrasonic aspirator can be particularly beneficial. These techniques are especially useful for anterior skull-base tumors and giant medial sphenoid-wing meningiomas.

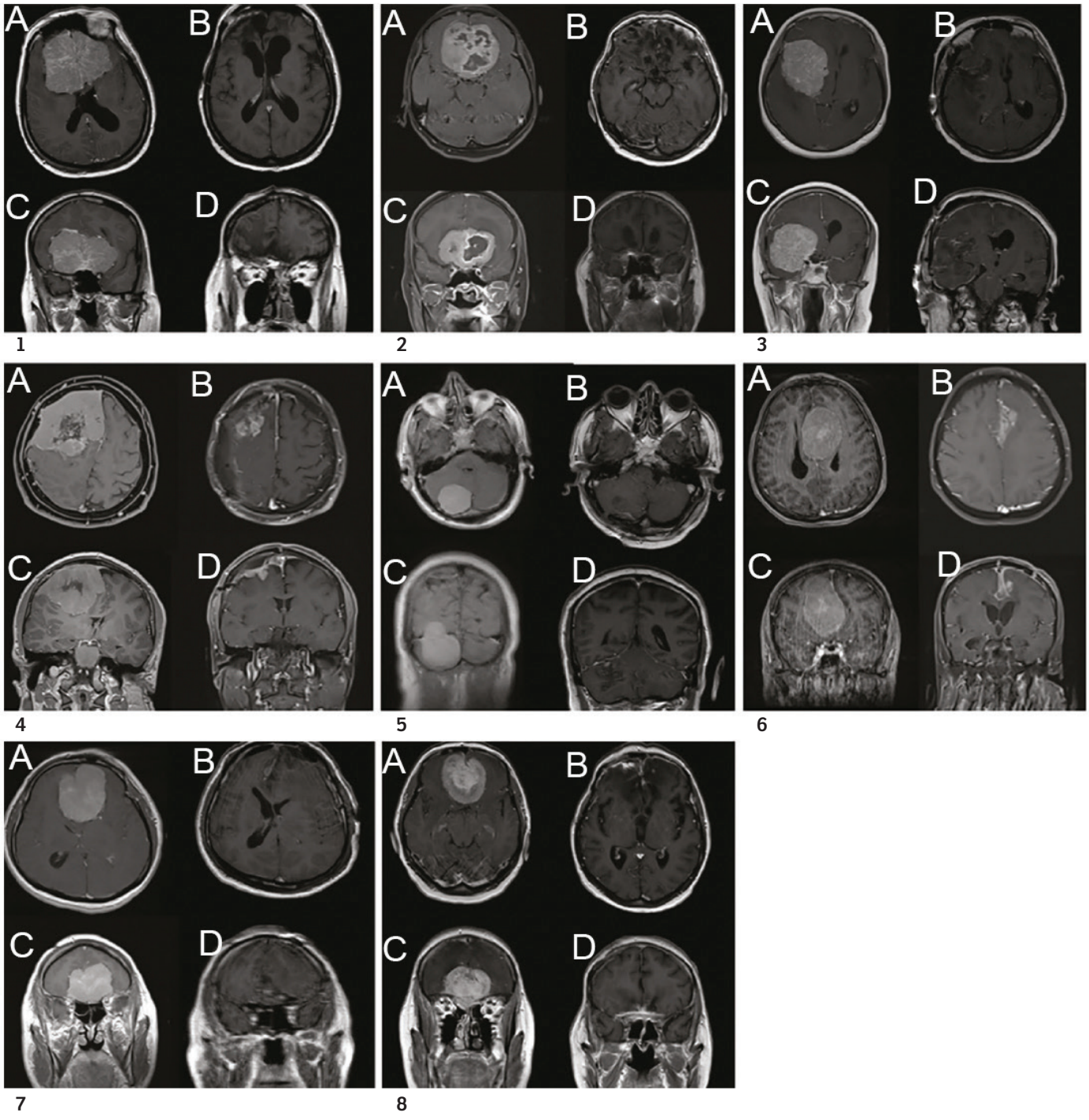


Figure 1-8. Pre- and post-operative contrast-enhanced MRI images of giant meningioma

A) Pre-operative axial images, B) Post-operative axial images, C) Pre-operative coronal images, D) Post-operative coronal images

MRI: Magnetic resonance imaging

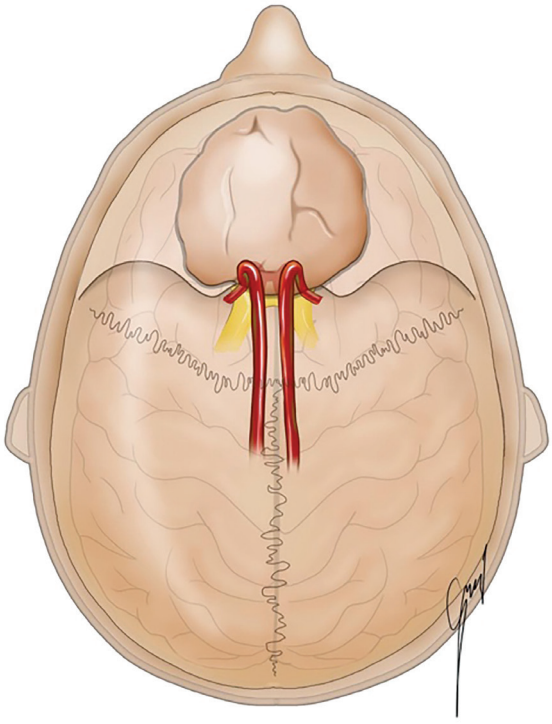


Figure 9. Illustration of a giant meningioma

Resection of giant petroclival meningiomas is demanding due to their size and the distortion of surrounding anatomical landmarks, paralleling the challenges encountered in clinoidal and olfactory groove lesions. For posterior fossa GIMs, Sanai described a modified far-lateral approach that enables gross total resection while maintaining low morbidity and mortality rates[18]. This technique provides several operative benefits, including enhanced tumor exposure, improved control of venous bleeding, and early entry into the spinal subarachnoid cisterns. Accurate identification of eloquent neural structures and major vascular elements, combined with meticulous manipulation of a tense and edematous brain, are critical to achieving maximal safe resection in cases of GIMs.

In their series, Chan and Thompson[15] reported that gross total resection was accomplished in 83% of patients, whereas subtotal removal was necessary in 17% of cases. Reported complication rates following surgical treatment of GIMs include postoperative neurological deterioration in 10.8% of patients, a perioperative mortality rate of 4%, and a tumor recurrence rate of 22%.

Excessive intraoperative hemorrhage remains a major operative challenge in GIM resections. The pronounced vascularity of these tumors may originate from multiple sources, including dural, pial, and transosseous arterial feeders, thereby increasing the risk of significant blood loss during dissection.

Preoperative endovascular embolization targeting the dominant arterial supply has been proposed as a strategy to mitigate intraoperative bleeding. Nevertheless, as emphasized by Bendszus et al.[19] effective reduction in blood loss is achieved only when embolization is complete, whereas partial devascularization offers limited benefit[20,21]. In our cohort, preoperative embolization was not performed, which likely contributed to the relatively elevated intraoperative blood loss. Giant meningiomas located at the olfactory groove or planum sphenoidale may derive their vascular supply from branches of the internal carotid artery, the middle meningeal artery, or the distal maxillary artery, thereby increasing the potential risk of visual compromise during surgical manipulation.

To mitigate this risk, Manjila et al.[22] described an extracranial approach involving open or endoscopic ligation of the ethmoidal arteries via transcaruncular or Lynch approaches prior to tumor resection. In our series, a comparable technique was used in patients with olfactory groove meningiomas.

Furthermore, the integration of advanced preoperative imaging modalities and intraoperative adjuncts-including cortical mapping, neuronavigation systems, and frameless stereotactic guidance-has substantially improved both operative safety and resection outcomes in the management of GIMs. Neuronavigation played a crucial role in achieving better resection outcomes, influencing overall survival. In this cohort, patients with skull base meningiomas had a better prognosis, likely due to their lower histological grade and improved resection status. In cases where the tumor involves critical regions, such as the petroclival area, selectively pursuing subtotal resection followed by radiotherapy may be a more reasonable approach than attempting gross total resection[10]. This is particularly true when gross resection poses significant risks to the patient's neurological function. The extent of tumor resection is typically assessed using the Simpson grading system, and in our series, we achieved gross total resection (Simpson grade 2) in six cases. Subtotal resection (Simpson grade 4) was performed in two patients because of the tumor's proximity to critical structures. Neuronavigation, in particular, allows surgeons to plan the shortest and safest route to the tumor and design an optimal craniotomy, thereby improving tumor

resection rates and patient survival. Complications during GIM surgeries are uncommon, but can include tumor bed hematoma, cerebrospinal fluid collection or leak, and infection. Inadequate hemostasis at the tumor bed can predispose patients to postoperative hematoma formation and associated neurological complications, whereas incomplete dural closure increases the risk of cerebrospinal fluid accumulation or leakage. Notably, in our series, no instances of hematoma or cerebrospinal fluid fistula were observed. These outcomes highlight the importance of meticulous surgical planning and technique to minimize complications in giant meningioma surgery.

Several factors can influence recurrence-free survival (RFS) in GIM cases. Incomplete resection (Simpson grade 4) and poor histological grade negatively impacted RFS.

A meticulous preoperative evaluation, thorough understanding of risk factors, an effective surgical approach, and judicious use of intraoperative adjuncts (such as neuronavigation) are critical to optimizing surgical outcomes. These elements collectively play a pivotal role in ensuring both safety and efficacy in the resection of GIMs. Giant meningiomas are extremely rare and pose significant challenges to total resection. GIMs exert a significant mass effect on adjacent brain tissue, leading to increased intracranial pressure, compression-related hemodynamic changes, and disruption of normal neural function. Vasogenic edema in the surrounding parenchyma is observed in over half of cases and can often be mitigated preoperatively through corticosteroid therapy.

Surgical management of GIMs presents multiple challenges. Prolonged operative times increase the risk of intraoperative blood loss and often necessitate careful dissection of encased cerebral arteries and cranial nerves. Tumor exposure may be limited by bridging veins and surrounding vasculature, making preoperative identification of critical neurovascular structures essential for safe resection. Initial operative steps typically involve coagulation of feeding vessels, with particular attention to avoiding injury to tense or edematous brain tissue throughout the procedure.

For skull base GIMs, cerebrospinal fluid drainage via transsylvian dissection facilitates brain relaxation, allowing internal debulking of the tumor. Endoscopic techniques, particularly for anterior skull base lesions, enable early devascularization and mobilization of the mass. Additional operative strategies that are critical for optimizing outcomes include proper patient positioning to reduce venous hypertension, selection of the most appropriate surgical corridor to minimize parenchymal injury, maintenance of excellent visualization of the operative field, and early identification of vital neurovascular structures. Meticulous dissection, preservation of bridging veins and perforating vessels, maintenance of the arachnoid plane, and diligent hemostasis are pivotal in reducing complications and improving postoperative outcomes.

Surgical planning must account for tumor location, size, and growth pattern. Convexity meningiomas typically require craniotomies larger than the lesion itself to allow safe manipulation, whereas skull base tumors demand a tailored approach. Craniofacial approaches may be indicated for extensive anterior skull base lesions, while pterional, subfrontal, or endoscopic endonasal routes are employed for tumors arising from the anterior fossa or tuberculum sellae. Intraventricular lesions can be accessed via tubular retractors with endoscopic assistance or through an interhemispheric transcallosal corridor.

The overarching goal of surgery is to maximize tumor removal while preserving neurological function and quality of life. Factors influencing the choice of surgical strategy include tumor location, direction of growth, involvement of adjacent neurovascular structures, patient age, and the operating surgeon's experience. The substantial size of GIMs can limit the extent of resection and be associated with higher recurrence rates, increased perioperative morbidity and mortality, and variable overall survival. Prognostic outcomes depend on multiple parameters, including tumor size and location, completeness of resection, histopathological grade, and use of adjuvant therapies. Notably, extensive vasogenic edema involving the bifrontal white matter in anterior skull-base meningiomas may predict poorer postoperative recovery.

Study Limitations

This study is limited by its retrospective design, small sample size, and relatively short and variable follow-up period. The heterogeneity of tumor locations limits meaningful subgroup comparisons, while the absence of inferential statistical analysis precludes definitive conclusions regarding prognostic factors and RFS. In addition, tumor volume was estimated using a simplified radiological formula rather than segmentation-based volumetry, and functional outcomes were not assessed using standardized scales. All procedures were performed by a single surgeon, which limits the generalizability of the findings despite reducing inter-operator variability.

Conclusion

GIM are rare and technically challenging lesions that require meticulous preoperative planning and individualized surgical strategies. Their large size, extensive vascularity, peritumoral edema, and frequent involvement or displacement of critical neurovascular structures substantially increase the risk of intraoperative blood loss, prolonged operative time, neurological morbidity, and incomplete resection. In our series, gross total resection was achieved in most patients, whereas subtotal resection was selected when tumor involvement of critical neurovascular structures made complete removal unsafe. Tumor location, extent of resection, and histopathological grade appeared to be important determinants of postoperative outcome and RFS. The use of neuronavigation, appropriate skull base or cranial approaches, early identification and control of feeding vessels, internal tumor debulking, cerebrospinal fluid drainage, and meticulous preservation of neurovascular structures may facilitate maximal safe resection. Given the rarity of GIMs and the limited number of patients in the present series, larger multicenter studies with longer follow-up are warranted to better define prognostic factors and establish standardized surgical strategies.

Ethics

Ethics Committee Approval: Permission for the study was obtained from the University of Health Sciences Türkiye, Bakırköy Dr. Sadi Konuk Training and Research Hospital Clinical Research Ethics Committee (approval no: 2025-01-35, date: 24.10.2025).

Informed Consent: Retrospective analysis.

Footnotes

Authorship Contributions

Concept: Y.D., M.E.G., E.G., Design: Y.D., M.E.G., Data Collection or Processing: Y.D., M.E.G., E.G., A.G.Ş., Literature Search: Y.D., M.E.G., E.G., Writing: Y.D., M.E.G., E.G., O.K.D., A.G.Ş., B.D., E.E., A.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.

References

1. Lamszus K. Meningioma pathology, genetics, and biology. J Neuropathol Exp Neurol. 2004;63:275-86.
2. Commins DL, Atkinson RD, Burnett ME. Review of meningioma histopathology. Neurosurg Focus. 2007;23:E3.
3. Narayan V, Bir SC, Mohammed N, Savardekar AR, Patra DP, Nanda A. Surgical management of giant intracranial meningioma: operative nuances, challenges, and outcome. World Neurosurg. 2018;110:e32-41.
4. Yaşar S, Kırık A. Surgical management of giant intracranial meningiomas. Eurasian J Med. 2021;53:73-8.
5. Nakasu S, Nakasu Y, Fukami T, Jito J, Nozaki K. Growth curve analysis of asymptomatic and symptomatic meningiomas. J Neurooncol. 2011;102:303-10.
6. da Silva CE, de Freitas PE. Large and giant skull base meningiomas: the role of radical surgical removal. Surg Neurol Int. 2015;6:113.
7. Behari S, Das KK, Kumar A, et al. Large/giant meningiomas of posterior third ventricular region: falcotentorial or velum interpositum? Neurol India. 2014;62:290-5.
8. Sanikommu S, Panchawagh S, Eatz T, et al. Recurrence of atypical and anaplastic intracranial Meningiomas: a meta-analysis of risk factors. Clin Neurol Neurosurg. 2024;244:108450.
9. Karımzada G, Evleksiz Karımzada D, Erol G, et al. Transorbital neuroendoscopic surgery for treatment of sphenoid wing meningiomas extending to the cavernous sinus: clinical implications and a technical illustration. Neurosurg Focus. 2024;56:E8.
10. Gurses ME, Zengin HY, Shikhaliyeva A, Askun CS, Mut M. Early postoperative adjuvant radiotherapy versus active monitoring after gross total resection for atypical meningiomas: factors associated with early recurrence. Turk Neurosurg. 2023;33:635-41.
11. Benz LS, Wrensch MR, Schildkraut JM, et al. Quality of life after surgery for intracranial meningioma. Cancer. 2018;124:161-6.
12. Bailo M, Gagliardi F, Boari N, Castellano A, Spina A, Mortini P. The role of surgery in meningiomas. Curr Treat Options Neurol. 2019;21:51.
13. Figarella-Branger D, Appay R, Metais A, et al. La classification de l'OMS 2021 des tumeurs du système nerveux central [The 2021 WHO classification of tumours of the central nervous system]. Ann Pathol. 2022;42:367-82. French.
14. Tuna M, Gocer AI, Gezercan Y, et al. Huge meningiomas: a review of 93 cases. Skull Base Surg. 1999;9:227-38.
15. Chan RC, Thompson GB. Morbidity, mortality, and quality of life following surgery for intracranial meningiomas. A retrospective study in 257 cases. J Neurosurg. 1984;60:52-60.
16. Wang C, Li P. Risk factors for intraoperative blood loss in resection of intracranial meningioma: analysis of 530 cases. PLoS One. 2023;18:e0291171.
17. Attia M, Umansky F, Paldor I, Dotan S, Shoshan Y, Spektor S. Giant anterior clinoidal meningiomas: surgical technique and outcomes. J Neurosurg. 2012;117:654-65.
18. Sanai N, McDermott MW. A modified far-lateral approach for large or giant meningiomas of the posterior fossa. J Neurosurg. 2010;112:907-12.
19. Bendszus M, Rao G, Burger R, et al. Is there a benefit of preoperative meningioma embolization? Neurosurgery. 2000;47:1306-11; discussion 1311-2.
20. Latchaw RE. Preoperative intracranial meningioma embolization: technical considerations affecting the risk-to-benefit ratio. AJNR Am J Neuroradiol. 1993;14:583-6.
21. Bendszus M, Rao G, Burger R, et al. Is there a benefit of preoperative meningioma embolization? Neurosurgery. 2000;47:1306-11.
22. Manjila S, Cox EM, Smith GA, et al. Extracranial ligation of ethmoidal arteries before resection of giant olfactory groove or planum sphenoidale meningiomas: 3 illustrative cases with a review of the literature on surgical techniques. Neurosurg Focus. 2013;35:E13.

DOI: 10.4274/anatoljmed.2026.09226
Anatol J Gen Med Res 2026;36(2):132-41

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Integrative Analysis of Stress During Academic Exams Using Wearable Sensor Data

Akademik Sınavlarda Stres Tepkilerinin Giyilebilir Sensör Verileriyle Bütüncül Analizi

Hasan Kazdağlı[1], Hasan Fehmi Özel[2]

[1]İzmir University of Economics Vocational School of Health Services, Department of Medical Services and Techniques, İzmir, Türkiye
[2]Manisa Celal Bayar University Vocational School of Health Services, Department of Medical Services and Techniques, Manisa, Türkiye

Cite as: Kazdağlı H, Özel HF. Integrative analysis of stress during academic exams using wearable sensor data. Anatol J Gen Med Res. 2026;36(2):132-41

Abstract

Objective: To comprehensively analyze students' physiological stress responses and their relationship with academic performance by using heart rate variability (HRV) and electrodermal activity (EDA) data collected via wearable sensors during academic examinations.

Methods: This study utilized the publicly available wearable exam stress dataset from PhysioNet. HRV and EDA data were recorded using Empatica E4 wristbands worn by ten university students during three exam sessions (midterm 1, midterm 2, and final). HRV parameters [low frequency (LF), high frequency (HF), LF/HF] and the tonic and phasic components of EDA were extracted. Differences between sessions were analyzed using the Mann-Whitney U test, while correlations with exam grades were evaluated using Spearman's rank correlation coefficient. In addition, individual variability and the relationship between HRV and EDA were examined on a session-by-session basis.

Results: Among HRV parameters, LF power (p=-0.440, p=0.015) and LF/HF ratio (p=-0.363, p=0.048) showed significant negative correlations with academic performance. Mean EDA values did not significantly differ across sessions (p>0.7); however, phasic components indicated increased arousal during the midterm 2 session. Correlations between EDA and grades were negative in midterm 1 (p=-0.552) and moderately positive in the final (p=0.406). Combined analysis of HRV and EDA data revealed concurrent autonomic dysregulation and acute stress responses, particularly during the midterm 2 session.

Conclusion: These findings indicate that academic examinations are associated with heterogeneous autonomic nervous system response patterns among students. The integrated evaluation of HRV and EDA contributes to a more comprehensive characterization of physiological stress responses at both the group and individual levels. Wearable sensor-derived physiological data may offer a useful framework for informing future research on stress monitoring and the development of individualized support approaches in academic settings.

Keywords: Heart rate variability, electrodermal activity, exam stress, integrative analysis

Öz

Amaç: Bu çalışmanın amacı, akademik sınavlar sırasında giyilebilir sensörlerle elde edilen kalp hızı değişkenliği (HRV) ve elektrodermal aktivite (EDA) verilerini kullanarak öğrencilerin fizyolojik stres tepkilerini ve bu tepkilerin akademik başarıyla ilişkisini bütüncül olarak analiz etmektir.

Yöntem: Bu çalışmada, PhysioNet'te açık erişimli olarak sunulan *wearable exam stress dataset* kullanıldı. On üniversite öğrencisinin üç sınav oturumu (ara sınav[1], ara sınav[2] ve final) sırasında giyilebilir Empatica E4 bilekliği ile kaydedilen HRV ve EDA verileri analiz edildi. HRV parametreleri [düşük frekans (LF),

Address for Correspondence/Yazışma Adresi: Asst. Prof., Hasan Kazdağlı, İzmir University of Economics Vocational School of Health Services, Department of Medical Services and Techniques, İzmir, Türkiye
E-mail: kazdaglihasan@gmail.com
ORCID ID: orcid.org/0000-0001-6617-604X

Received/Geliş tarihi: 27.11.2025
Accepted/Kabul tarihi: 29.01.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

yüksek frekans (HF), LF/HF] ve EDA'nın tonik-fazik bileşenleri çıkarıldı; oturumlar arası farklar Mann-Whitney U testi ile, notlarla ilişkiler ise Spearman korelasyonu ile değerlendirildi. Ayrıca bireysel değişkenlik ve HRV-EDA ilişkileri de oturum bazında incelendi.

Bulgular: HRV parametrelerinden LF gücü (p=-0,440, p=0,015) ve LF/HF oranı (p=-0,363, p=0,048) akademik başarıyla anlamlı düzeyde negatif korelasyon gösterdi. EDA ortalama değerleri oturumlar arasında anlamlı farklılık göstermedi (p>0,7); ancak fazik bileşenler ara sınav 2 oturumunda artmış uyarılma düzeylerini yansıttı. EDA ile notlar arasında ara sınav 1'de negatif (p=-0,552) ve final oturumunda pozitif yönde orta düzeyde (p=0,406) korelasyon gözlendi. HRV ve EDA verilerinin birlikte incelenmesi, özellikle ara sınav 2'de otonom düzensizlik ile ani stres tepkilerinin birlikte seyrettiğini gösterdi.

Sonuç: Bu bulgular, akademik sınavların öğrencilerde heterojen otonom sinir sistemi yanıtlarıyla ilişkili olduğunu göstermektedir. HRV ve EDA'nın bütüncül olarak değerlendirilmesi, stres tepkilerinin hem grup hem de bireysel düzeyde daha kapsamlı biçimde tanımlanmasına katkı sağlamaktadır. Giyilebilir sensörlerden elde edilen fizyolojik veriler, akademik ortamlarda stres izlenmesine yönelik gelecekteki araştırmalar ve bireyselleştirilmiş destek yaklaşımlarının geliştirilmesi için yararlı bir çerçeve sunabilir.

Anahtar Kelimeler: Kalp atım hızı değişkenliği, elektrodermal aktivite, sınav stresi, bütüncül analiz

Introduction

Exam stress is common among students at various academic levels and typically results from a complex interaction of psychological, physiological, and behavioral factors[1]. Understanding the impact of exam stress on students requires a multifaceted approach, as it significantly influences both their mental well-being and academic performance[2]. Although exam stress is considered a natural part of academic life, it is closely linked to students' stress perceptions, coping strategies, and ultimately their academic achievement[3].

The impact of stress on academic achievement is a topic that has been examined by numerous researchers. The stress experienced by students leads to negative outcomes such as decreased exam performance and reduced course attendance[4]. Similarly, students who experience high levels of stress often suffer from fatigue, which creates a cycle between declining performance and increasing stress[5]. The widespread effects of stress are not limited to academic outcomes; by influencing behavior, it can lead to maladaptive coping mechanisms such as substance use and social withdrawal, which further exacerbate academic difficulties[6,7].

Students who successfully develop effective ways of coping with stress tend to show improved academic performance, along with reduced habits such as substance and tobacco use[8]. Therefore, early identification of students who are prone to stress and providing them with effective stress-management skills are of great importance[9]. Today, the most commonly used method for assessing exam stress consists of questionnaire-based studies in which students evaluate themselves[10]. Many studies have shown that these scales may present limitations in terms of reliability and validity, which raises questions about their usefulness in accurately measuring exam-related stress[11].

First, the development of stress measurement scales often involves multiple psychological constructs, which can lead to inconsistencies in practice. For example, although the beliefs about stress scale has demonstrated adequate validity when administered before and after major exams, the reliability of such scales may vary depending on contextual factors such as the demographic characteristics of the student sample[12]. This indicates that although some scales may provide reliable results under specific conditions, they are not universally applicable across different academic settings or types of stress, which weakens overall assessment reliability.

Moreover, Zunhammer et al.[13] have shown that certain psychological traits can significantly distort the measurements obtained from stress scales administered during examinations. This study demonstrated that some personality characteristics are not sufficiently accounted for in traditional stress assessment models, which may negatively affect the reliability and interpretability of the results. The presence of heightened somatic symptoms associated with exam stress, alongside psychological factors, makes the direct interpretation of scores obtained from stress measurement tools more complex and raises concerns about their reliability across different student groups[14].

Research shows that academic stress significantly affects various physiological markers, including cortisol levels, heart rate variability (HRV), and electrodermal activity (EDA)[3,15].

HRV is widely used as a valuable biomarker for evaluating stress responses, particularly in high-pressure environments

such as exams[16]. The relationship between HRV and academic stress has been examined in various studies. For example, Obi et al.[17] reported that students exhibited lower HRV values before and during exams, which was interpreted as an indicator of heightened mental stress. This reduced HRV observed during exam periods has been associated with increased sympathetic dominance and decreased parasympathetic activity. Similarly, Hammoud et al.[18] reported a decrease in students' HRV levels during exams, with a significant increase observed after exams. This finding indicates a reduction in stress following the exam period.

EDA, also commonly known as galvanic skin response, is an important physiological measure that reflects sympathetic nervous system activity[19]. This measurement has various applications in evaluating stress, particularly in educational or competitive environments such as exams[19]. EDA stands out as an effective and non-invasive indicator of stress because it correlates with sympathetic nervous system arousal, which varies with stress levels[20]. EDA measurements can evaluate emotional and stress-related states without requiring complex equipment, and this method is considered ideal for educational settings where exam stress is common[21]. Furthermore, studies have shown that stress can lead to a decline in cognitive performance and an increase in anxiety levels. This is particularly evident in high-pressure situations such as exams, and physiological responses like EDA can provide real-time feedback on students' stress levels[22,23].

The study conducted by Rahma et al.[24] demonstrated that both the tonic (baseline) and phasic (stimulus-responsive) components of EDA are effective in distinguishing different levels of cognitive and emotional stress. This finding indicates that EDA enables the monitoring of stress responses during critical moments such as exams[24].

However, studies that analyze EDA and HRV data together and relate them to students' academic performance remain limited. Therefore, in our study we aimed to examine the relationship between physiological stress responses and academic performance by comprehensively evaluating EDA and HRV data collected with wearable sensors during academic exams.

Accordingly, the present study adopts an exploratory and hypothesis-generating approach, aiming to characterize integrated autonomic stress responses during real-world academic examinations rather than to establish definitive causal relationships. By leveraging an ecologically valid wearable dataset, this work seeks to identify physiological patterns that may inform future large-scale and controlled investigations in educational psychophysiology.

Materials and Methods

Study Design and Dataset

This study was conducted using the "wearable exam stress dataset", an open-access data set[25]. The dataset contains electrophysiological signals collected from ten university students during three different academic exam sessions (midterm 1, midterm 2, and final), using the Empatica E4 wearable wrist device[26]. Heartbeat intervals [inter-beat interval (IBI)], EDA, and body temperature data were continuously recorded from participants using the devices.

The dataset does not include resting or non-exam baseline physiological recordings. Therefore, all analyses were conducted using within-session and between-session comparisons across examination periods, rather than deviations from individual baseline autonomic states.

Ethical Aspect of the Research

Our study was conducted with the approval of the İzmir University of Economics Health Sciences Research Ethics Committee (approval no: E-97429853-050.04-111205, date: 05.11.2025). The dataset is available to researchers via the PhysioNet platform, and because participant information is anonymized, an informed consent form is not required. In our study, no direct intervention or interaction was carried out, as the dataset was used solely for analysis. Therefore, the research qualifies as a secondary data analysis.

Participants

Data from 10 university students in the wearable exam stress dataset were analyzed. The participants were healthy individuals with no psychiatric or neurological diagnoses and were included in the study on a voluntary basis[25]. In the dataset, participants' identities were kept confidential, and only limited demographic information, such as age, sex, and academic performance, was provided. Participants included male and female students, and the reported age range was 20-25 years[25]. For each participant, electrophysiological data from three exam sessions (midterm 1, midterm 2, and final) and the students' exam scores were obtained.

Data Collection

Data were collected using the Empatica E4 wrist-worn device during participants' exam sessions. This wearable sensor can

simultaneously record various physiological parameters. In addition to the two primary electrophysiological signals, IBI and EDA, body temperature was also recorded within the scope of the study[24]. Figure 1 illustrates a representative IBI tachogram derived from wearable photoplethysmography (PPG) recordings.

IBI data obtained using an optical PPG sensor were used for HRV calculations. EDA data were continuously recorded using two electrodes placed on the inner surface of the wrist and evaluated as indicators of sympathetic nervous system activity. Body temperature measurements were continuously monitored via the temperature sensor located on the device's body and were analyzed for the time periods corresponding to the exam sessions[24]. Accordingly, skin temperature should be viewed as a complementary, context-dependent marker that gains interpretive value primarily when integrated with HRV and EDA indices.

For each participant, nearly 24 hours of data were recorded throughout the exam day[24]. Because of the time-stamped nature of the data, segments corresponding to the periods immediately before and during the exam were extracted and preprocessed for analysis. Noise, motion artifacts, and missing signals were removed using appropriate filtering and cleaning methods prior to data analysis.

Although the Empatica E4 device provides validated PPG-derived IBI measurements, it is acknowledged that wrist-based PPG signals are more susceptible to motion artifacts and may exhibit reduced precision compared to electrocardiography-based recordings. Consequently, HRV results should be interpreted with appropriate caution, particularly in high-movement contexts.

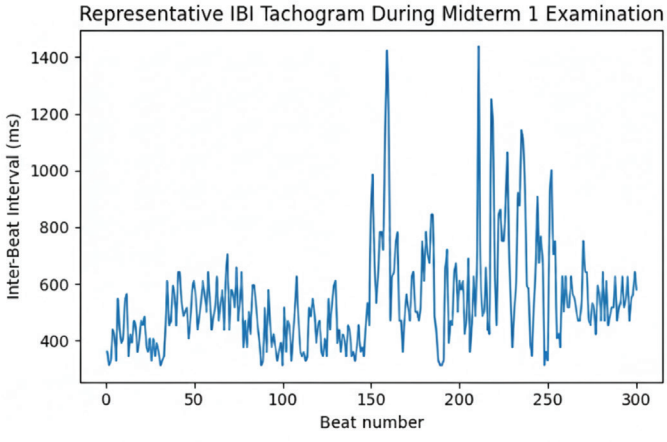


Figure 1. Representative inter-beat interval (IBI) tachogram derived from wearable photoplethysmography recordings during the midterm 1 examination. The tachogram, after artifact cleaning and preprocessing, illustrates beat-to-beat variability and demonstrates the signal quality used for subsequent HRV and non-linear analyses

HRV: Heart rate variability

HRV and EDA Analyses

IBI data obtained from the PPG sensor of the Empatica E4 device were used for HRV analysis. Raw IBI series were initially subjected to automated artifact detection and cleaning procedures to identify ectopic beats, motion-related disturbances, and missing values. Artifact-contaminated intervals were corrected using interpolation following sequence verification, and only segments with adequate signal quality and a minimum duration of 5 minutes were retained for subsequent analyses. Because the dataset did not include resting or non-exam baseline recordings, HRV metrics were interpreted using within-session and between-session comparisons across examination periods, rather than as deviations from individual normative autonomic values.

HRV analysis included time-domain, frequency-domain, and non-linear parameters. For frequency-domain analyses, low frequency [(LF): 0.04-0.15 Hz] and high frequency [(HF): 0.15-0.4 Hz] components, as well as the LF/HF ratio, were calculated[3]. These indices reflect sympathetic modulation, parasympathetic activity, and overall autonomic balance, respectively[27]. For frequency analysis, power spectral density was estimated using the Welch method[28]. HRV analysis was performed using the open-source Python libraries NeuroKit2 and HRV-analysis. HRV metrics were computed using open-source Python libraries NeuroKit2 and HRV-analysis, employing consistent preprocessing and parameter settings across all exam sessions to ensure methodological standardization. HRV parameters were calculated separately for each examination session and subsequently evaluated both across individuals and between exam types. In addition to conventional HRV indices, non-linear measures derived from detrended fluctuation analysis (DFA) were included. Specifically, the dS1, dS2, and TDs parameters, proposed by Ozel and Kazdagli[29] as quantitative indices of short-term, long-term, and total physiological complexity loss in heart rate time series, were computed to characterize stress-related alterations in autonomic complexity. Given the known limitations of wrist-based PPG-derived HRV measurements, particularly their susceptibility to motion artifacts, all results were interpreted with appropriate caution.

EDA signals were analyzed for the time intervals corresponding to each examination session. Raw EDA signals, sampled at 4 Hz, were initially inspected for signal dropouts and motion-related artifacts. Noise reduction was performed using low-pass filtering, and segments with excessive artifacts were excluded to ensure signal reliability. Because environmental and individual factors influencing EDA could not be experimentally controlled, all analyses were conducted using a consistent preprocessing pipeline across sessions to maintain methodological standardization.

Following preprocessing, EDA signals were decomposed into tonic (skin conductance level) and phasic [skin conductance response, (SCR)] components. The tonic component reflects the overall level of sympathetic arousal, whereas the phasic component captures rapid, stimulus-related sympathetic activations[24]. Decomposition was performed using the convex-optimization-based cvxEDA algorithm with default regularization parameters as implemented in NeuroKit2.

For each participant and examination session, summary metrics including mean EDA level, number of SCRs, and maximum SCR amplitude were calculated. All EDA analyses were conducted in Python using the NeuroKit2 and BioSPPy libraries, with identical parameter settings applied across sessions to ensure analytical consistency.

Academic Performance Data

To evaluate the participants' academic performance, individual exam scores from each session (midterm 1, midterm 2, and final) were used. These scores were provided to researchers as part of the "wearable exam stress dataset", with midterm 1 and midterm 2 graded on a 100-point scale. Final exam scores were normalized to a 100-point scale before being used in the analyses.

Statistical Analysis

Statistical analysis of the data was performed using GraphPad Prism (version 10.4.0). Data were expressed as medians. A significance level of $p<0.05$ was accepted for all analyses. To evaluate differences in HRV and EDA parameters across exam sessions (midterm 1, midterm 2, and final), the non-parametric Mann-Whitney U test was applied.

Spearman's rank correlation coefficient (ρ) was calculated to determine the relationships between HRV and EDA parameters and exam scores. In addition, individual physiological response profiles were examined, and the relationships between HRV and EDA components were evaluated on a session-by-session basis.

Correlation plots present the correlation coefficients and their significance levels.

Exam score data were treated as continuous variables and analyzed at both the individual and session levels. Because the score distributions were not normal, non-parametric correlation analyses were preferred. The relationships between physiological data and academic performance were examined using Spearman's rank correlation. Moreover, to assess the impact of differing stress levels on performance, participants were grouped according to HRV and EDA parameters.

Differences in EDA and HRV parameters across the different exams were assessed using the Friedman test. For parameters that reached statistical significance ($p<0.05$), post-hoc analyses (Wilcoxon matched-pairs test with Bonferroni correction) were performed.

Results

Comparison of HRV, EDA, and Body Temperature Parameters Across Exams

The HRV and EDA metrics, together with the corresponding examination scores, are summarized in Table 1 and presented as median ± interquartile range (IQR). Both DFAα_1 and the stress index exhibited pronounced elevations during midterm 1 ($p=0.000265$ and $p=0.00145$, respectively), indicating an early-session perturbation in autonomic regulatory dynamics. Conversely, the tonic component of the EDA signal was significantly attenuated during the final examination ($p=0.00048$), suggesting diminished sympathetic arousal in later stages of the assessment sequence. No additional physiological parameters demonstrated session-dependent divergence ($p>0.05$), indicating relative stability in other autonomic indices across the exam period.

Within-participant analyses provided further evidence of session-specific modulation of autonomic function. Mean heart rate (Figure 2) reached its maximum during midterm 1 and subsequently declined, while DFAα_1 (Figure 3), a well-established marker of short-term fractal cardiac dynamics, demonstrated an analogous pattern of early-session elevation followed by progressive reduction. Friedman test results corroborated these effects (mean hazard ratio: $p=0.0001$; DFAα_1: $p=0.0498$). Post-hoc contrasts revealed significant decreases in mean heart rate between midterm 1 and midterm 2, and between midterm 1 and the final examination ($p=0.0117$ for both). A significant decline in DFAα_1 emerged exclusively between midterm 1 and the final

Table 1. HRV parameters, EDA values, and exam scores of students across the exams

Parameters	Midterm 1	Midterm 2	Final
Exam score (%)	77.50±14.75	77.00±22.50	76.50±19.62
Mean HR (bpm)	121.65±11.00	95.83±11.12	91.41±15.45
SDNN (ms)	30.35±25.81	31.88±13.06	27.38±7.55
RMSSD (ms)	16.65±10.77	22.71±7.78	21.86±1.83
Stress index	18.48±8.89	10.90±4.09	13.35±3.06
VLF (%)	74.14±27.94	69.13±15.06	71.85±10.16
LF (%)	45.41±21.66	46.28±12.34	50.95±10.18
HF (%)	36.51±31.54	43.19±11.42	40.65±8.49
LF/HF ratio	1.49±2.47	1.07±0.33	1.25±0.50
$DFA\alpha_1$	1.43±0.28	1.20±0.18	1.17±0.14
$DFA\alpha_2$	0.70±0.35	0.61±0.30	0.54±0.06
Mean EDA (μS)	0.219±0.171	0.234±0.224	0.191±0.379
Tonic component (μS)	0.153±0.123	0.134±0.182	0.026±0.420
Phasic component (μS)	0.296±0.312	0.305±0.221	0.266±0.189

Data are presented as median ± interquartile range

HRV: Heart rate variability, EDA: Electrodermal activity, HR: Heart rate, SDNN: Standard deviation of normal-to-normal intervals, RMSSD: Root mean square of successive differences, VLF: Very-low frequency, LF: Low frequency, HF: High frequency, $DFA\alpha_1$: Short-term detrended fluctuation analysis, $DFA\alpha_2$: Long-term detrended fluctuation analysis

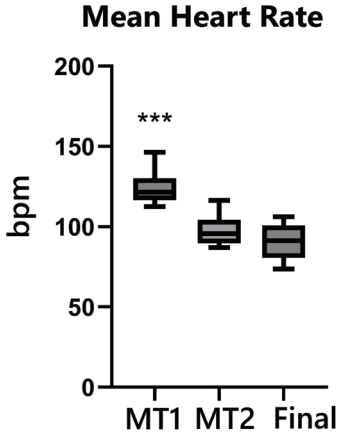


Figure 2. Comparison of mean heart rate across exams. Friedman test followed by Bonferroni-corrected Wilcoxon post-hoc test

****: p<0.001 MT2 and final*

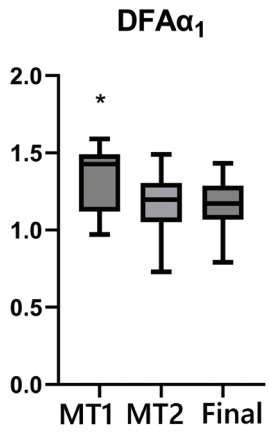


Figure 3. Comparison of $DFA\alpha_1$ across exams

*Friedman test, *: p < 0.05 MT1 vs. final*

(p=0.0352, Bonferroni-corrected), further supporting the notion of a front-loaded autonomic load.

Non-linear HRV descriptors yielded additional insights into the temporal structure of cardiac variability under academic stress. Both dS1 and TdS, indices reflecting short-term and global reductions in physiological complexity, were markedly elevated during midterm 1 (p=0.0326 and p=0.0211, respectively). Median ± IQR distributions for dS1 were 0.0886±0.1238 (midterm 1), 0.0448±0.0403 (midterm 2), and 0.0792±0.0565 (final); for dS2, 0.0157±0.0406, 0.0188±0.0331, and 0.0243±0.0188; and for TdS, 0.0886±0.1238, 0.0448±0.0403, and 0.0792±0.0565 across the three exam sessions (Figure 4). Collectively, these signatures point to a transient disruption of physiological complexity specifically at the onset of the examination sequence.

Thermoregulatory indices exhibited a comparable pattern of session-related fluctuations. Median ± IQR values for body temperature were 26.58±1.40 °C (midterm 1), 27.50±1.94 °C (midterm 2), and 27.62±3.21 °C (final), indicating modest yet structured variation across assessment periods (Figure 5). These findings underscore the sensitivity of peripheral temperature to examination-induced psychophysiological load and provide an ancillary lens through which to contextualize autonomic reactivity.

Taken together, these results delineate a coherent physiological profile in which early-session examinations elicit heightened sympathetic signatures and reduced autonomic complexity, with subsequent attenuation of these signatures across the exam trajectory. Such temporal dynamics align with theoretical models of anticipatory stress and habituation within real-world cognitive performance contexts.

Relationship Between Exam Scores and Physiological Parameters

Table 2 delineates the associations between exam performance and electrophysiological stress markers derived from HRV and EDA. Correlation analyses demonstrated a statistically significant inverse association between LF power, reflecting baroreflex-mediated sympathetic vasomotor modulation, and exam scores (ρ=-0.440, p=0.015). Likewise, the LF/HF ratio, a canonical index of sympathovagal balance, exhibited a significant negative correlation with performance (ρ=-0.363, p=0.048), indicating that a shift toward sympathetic predominance was linked to reduced academic outcomes.

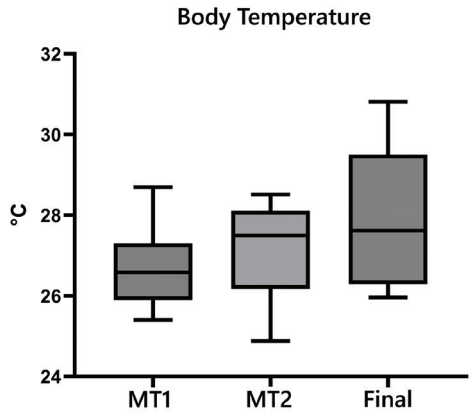


Figure 5. Comparison of body temperature across exams. Values represent wrist skin temperature recorded by the Empatica E4, which may be environmental conditions. Peripheral skin temperature did not differ significantly across examination sessions (Friedman test, p>0.05). Therefore, observed variations should be interpreted cautiously because of potential environmental influences

Table 2. Relationship between exam scores and stress markers

Parameters	Correlation coefficient (ρ)	p-value
LF power	-0.440	0.015
LF/HF ratio	-0.363	0.048
DFAα1	0.41	0.049

LF: Low frequency, HF: High frequency, DFAα1: Short-term detrended fluctuation analysis

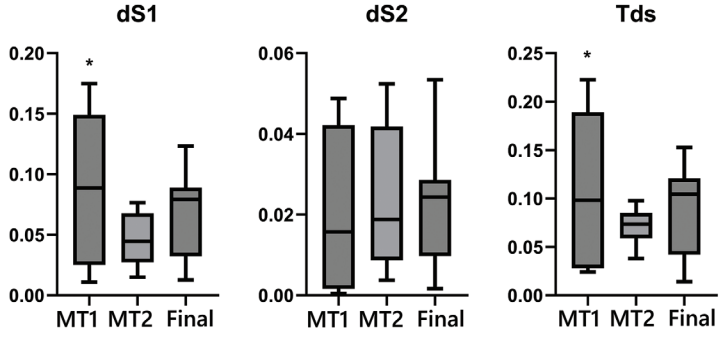


Figure 4. Comparison of physiological complexity loss parameters across exams

dS1: Short-term physiological complexity loss parameter derived from DFAα1, dS2: Long-term physiological complexity loss parameter derived from DFAα2, Tds: Total physiological complexity loss. Statistical analysis was performed using the Friedman test to assess differences across exam sessions (MT1, MT2, final). For parameters reaching statistical significance, post-hoc pairwise comparisons were conducted using the Wilcoxon matched-pairs test with Bonferroni correction. Asterisks indicate statistically significant differences relative to other exam sessions (: p<0.05)*

Conversely, DFA$\alpha_{1,}$ a non-linear metric capturing short-term fractal scaling properties and reflecting the system's intrinsic autonomic complexity, showed a significant positive correlation with exam scores (ρ=0.410, p=0.049). This pattern suggests that the preservation of fractal-like heartbeat dynamics and reduced autonomic rigidity may confer a performance advantage under evaluative stress.

Discussion

Analysis of electrophysiological data collected from 10 university students during three examination sessions revealed session-dependent variations in autonomic indices. Several HRV measures, including DFAα_1 and the HRV-derived stress index, differed across sessions, suggesting that cardiac autonomic dynamics are sensitive to examination context. However, given the exploratory design, a small and homogeneous sample, and the absence of baseline physiological recordings, these findings should be interpreted as descriptive associations rather than evidence of stress-induced physiological effects.

These observations are broadly consistent with prior studies conducted under naturalistic examination conditions. Dimitriev et al.[30] reported reductions in HRV complexity and elevations in DFAα_1 during examinations, with DFAα_1 associated with state anxiety rather than performance outcomes. Similarly, Kazdagli et al.[3] observed elevated DFAα_1 values in students with higher exam-related anxiety, alongside higher exam scores in some individuals, indicating that DFAα_1 does not uniformly reflect maladaptive autonomic dysregulation but may coexist with effective cognitive engagement under evaluative stress. Thus, DFAα_1 appears to function as a stress-responsive marker whose relationship with performance is context-dependent and moderated by individual stress-regulation capacity.

Consistent with this interpretation, DFAα_1 showed a positive association with exam scores, whereas LF power and the LF/HF ratio were negatively correlated with performance, linking sympathetic dominance to poorer academic outcomes[3]. This pattern aligns with a non-linear arousa-performance relationship, whereby moderate autonomic activation may support cognitive efficiency while excessive sympathetic dominance impairs it, in line with the Yerkes-Dodson framework[31]. As emphasized in previous reviews, such physiological-performance associations are typically modest and influenced by interindividual variability[32].

Among non-linear HRV metrics, TdS was elevated during the initial examination, suggesting greater disruption of autonomic complexity during early evaluative stress, consistent with prior reports of physiological complexity loss under stress[29,33]. The subsequent reduction in TdS across later sessions may reflect partial habituation; however, interpretation is constrained by the lack of baseline recordings and the use of wrist-based PPG-derived HRV signals, which are susceptible to motion artifacts.

EDA findings further supported sensitivity to exam-related arousal. As a peripheral index of sympathetic activation, EDA reflects emotional engagement and cognitive load[19,24] and higher tonic EDA levels have been reported in students with elevated test anxiety[3]. Nevertheless, EDA represents a nonspecific arousal signal influenced by environmental and individual factors[34,35], exhibits delayed responses relative to sympathetic activation[36], and lacks full methodological standardization across studies[37]. Accordingly, its interpretive value is greatest when integrated with complementary autonomic measures such as HRV[38-41].

Skin temperature provided additional, context-dependent information. Reductions observed in some participants are consistent with sympathetic vasoconstriction; however, substantial interindividual variability and sensitivity to environmental conditions limit its utility as a standalone marker of autonomic arousal[33].

Overall, this study provides an integrative, exploratory assessment of exam-related physiological patterns using wearable-derived measures of HRV, EDA, and skin temperature. While session-dependent autonomic variations and associations with academic performance were observed, conclusions are limited by sample size, the absence of baseline and subjective stress measures, environmental confounding, and reliance on offline signal processing. Future studies incorporating larger and more diverse cohorts, baseline-controlled and longitudinal designs, multimodal psychological–physiological assessments, and advanced analytic approaches, including artificial intelligence-based methods, will be essential to refine the use of wearable technologies for monitoring and managing exam-related stress.

Study Limitations

Several limitations of the present study should be acknowledged. First, the study relies on a secondary analysis of a publicly available dataset with a relatively small sample size ($n=10$) consisting of a single student cohort, which constrains the statistical power and the generalizability of the findings across broader and more diverse student populations. Second, the dataset lacked non-exam resting baseline physiological recordings; therefore, all autonomic evaluations were restricted to within-session and between-session comparisons rather than deviations from true resting autonomic states. Third, concurrent psychological assessments (such as validated self-report scales for state anxiety, test anxiety, or subjective stress perception) were not available, limiting the direct triangulation between subjective psychological experiences and physiological markers. Fourth, physiological data were acquired exclusively through wrist-worn wearable PPG and dry electrodermal electrodes (Empatica E4). Although ecologically valid and non-invasive, wrist PPG is inherently more susceptible to motion artifacts and subtle peak-detection errors than standard ECG, and peripheral skin conductance and temperature are subject to uncontrolled ambient and environmental factors. Finally, the current analyses were performed post-hoc and offline; real-time stress detection and continuous longitudinal monitoring across multiple academic periods remain to be implemented in future research.

Conclusion

In conclusion, this study demonstrates that academic examinations elicit dynamic, session-dependent autonomic nervous system responses that can be non-invasively tracked using wearable multi-sensor technology. Our findings indicate that heightened sympathetic dominance (reflected by elevated LF power and LF/HF ratio) is negatively associated with academic performance, whereas the preservation of cardiac autonomic complexity and short-term fractal dynamics (DFAα1) is positively correlated with exam achievement. Furthermore, the temporal trajectory of physiological metrics, characterized by peak cardiovascular arousal and physiological complexity loss during early exam sessions followed by subsequent attenuation, supports models of anticipatory stress and subsequent adaptation in evaluative academic environments. Combining HRV, EDA components, and peripheral temperature provides a multidimensional psychophysiological profile that surpasses single-parameter assessments. Despite the exploratory nature of this work, these results highlight the potential of wearable biosensors as objective tools for educational psychophysiology. Future research employing larger cohorts, controlled baseline protocols, multimodal psychological-physiological tracking, and real-time artificial intelligence-assisted analytics will be instrumental in developing individualized stress-monitoring frameworks and adaptive pedagogical interventions to enhance student well-being and academic performance.

Ethics

Ethics Committee Approval: Our study was conducted with the approval of the İzmir University of Economics Health Sciences Research Ethics Committee (approval no: E-97429853-050.04-111205, date: 05.11.2025).

Informed Consent: The dataset is available to researchers via the PhysioNet platform, and because participant information is anonymized, an informed consent form is not required.

Footnotes

Authorship Contributions

Design: H.K., Data Collection or Processing: H.K., H.F.Ö., Analysis or Interpretation: H.K., H.F.Ö., Literature Search: H.K., H.F.Ö., Writing: H.K., H.F.Ö.

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

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

References

1. Crego A, Carrillo-Diaz M, Armfield JM, Romero M. Stress and academic performance in dental students: the role of coping strategies and examination-related self-efficacy. J Dent Educ. 2016;80:165-72.
2. Beltrán-Velasco AI, Bellido-Esteban A, Ruisoto-Palomera P, Clemente-Suárez VJ. Use of portable digital devices to analyze autonomic stress response in psychology objective structured clinical examination. J Med Syst. 2018;42:35.
3. Kazdagli H, Ozel HF, Ozturk S, et al. Electrophysiological detection of exam stress in health schools' students. Physiol Int. 2024;111:222-41.
4. Akanpaadgi E, Binpimbu F, Kuuyelleh EN. The impact of stress on students' academic performance. Eureka J Educ Res. 2023;2:60-7.
5. Lee SM, Oh Y. The mediator role of perceived stress in the relationship between academic stress and depressive symptoms among e-learning students with visual impairments. Journal of Visual Impairment & Blindness. 2017;111:123-34.
6. Shrivastava U, Thakur A, Singh TP, et al. Comparative analysis of stress levels among undergraduate students at St. John's College, Agra. Cureus. 2024;16:e67452.

7. Rosaline MD, Anggraeni DT. Factors related to academic stress during the covid-19 pandemic in nursing students of UPN Veteran Jakarta. Advances in Health Sciences Research.2020;30:374-9.
8. Joseph N, Nallapati A, Machado MX, et al. Assessment of academic stress and its coping mechanisms among medical undergraduate students in a large Midwestern University. Curr Psychol. 2021;40:2599-609.
9. Nagle YK, Sharma U. Academic stress and coping mechanism among students: an Indian perspective. J Child Adolesc Psychol. 2018;2:6-8.
10. Bedewy D, Gabriel A. Examining perceptions of academic stress and its sources among university students: the perception of academic stress scale. Health Psychol Open. 2015;2:2055102915596714.
11. Razavi T. Self-report measures: an overview of concerns and limitations of questionnaire use in occupational stress research. 2001:1-29.
12. Laferton JA, Stenzel NM, Fischer S. The Beliefs about stress scale (BASS): development, reliability, and validity. Int J Stress Manag. 2018;25:72.
13. Zunhammer M, Eberle H, Eichhammer P, Busch V. Somatic symptoms evoked by exam stress in university students: the role of alexithymia, neuroticism, anxiety and depression. PLoS One. 2013;8:e84911.
14. Danthony S, Mascret N, Cury F. Development and validation of a scale assessing test anxiety in physical education. J Teach Phys Educ. 2019;38:357-66.
15. Murphy L, Denis R, Ward CP, Tartar JL. Academic stress differentially influences perceived stress, salivary cortisol, and immunoglobulin-a in undergraduate students. Stress. 2010;13:366-71.
16. Guimarães T, Silva AP, Monteiro I, Gomes R. Stress in university students and cardiovascular response to academic stressors. 2014.
17. Obi A, Nwobodo EO, Dimkpa U, Maduka SO, Fintan E. Assessment of stress level of young undergraduates before and after a degree examination using heart rate variability analysis. European Journal of Medical and Health Sciences. 2021;3:1-6.
18. Hammoud S, Karam R, Mourad R, Saad I, Kurdi M. Stress and heart rate variability during university final examination among lebanese students. Behavioral Sciences. 2019;9:3.
19. Pop-Jordanova N, Pop-Jordanov J. Electrodermal activity and stress assessment. Pril (Makedon Akad Nauk Umet Odd Med Nauki). 2020;41:5-15.
20. Sikder MSBB, Salimullah SM, Ahmed I, Hossain D, Rajib MA, Wadud K. Galvanic skin response based cognitive load measurement for online exams. IEEE. 2021.
21. Horvers A, Tombeng N, Bosse T, Lazonder AW, Molenaar I. Detecting emotions through electrodermal activity in learning contexts: a systematic review. Sensors (Basel). 2021;21:7869.
22. Al Jammali Z, Al Yasiry A, Alsaadi MA, Alyassery AAA. The impact of anxiety on undergraduate students' outcomes during prosthetic exam. J Stomatol. 2024;77:29-33.
23. Roos AL, Goetz T, Krannich M, et al. Control, anxiety and test performance: self-reported and physiological indicators of anxiety as mediators. Br J Educ Psychol. 2023;93:72-89.
24. Rahma ON, Putra AP, Rahmatillah A, et al. Electrodermal activity for measuring cognitive and emotional stress level. J Med Signals Sens. 2022;12:155-62.
25. Amin R, Wickramasuriya D, Faghih RT. A wearable exam stress dataset for predicting cognitive performance in real-world settings (version 1.0.0). PhysioNet. 2022.
26. Amin R, Wickramasuriya DS, Faghih RT. A wearable exam stress dataset for predicting grades using physiological signals. IEEE. 2022.
27. Shaffer F, Ginsberg JP. An overview of heart rate variability metrics and norms. Front Public Health. 2017;5:258.
28. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. Circulation. 1996;17:354-81.
29. Ozel HF, Kazdagli H. A simple approach to determine loss of physiological complexity in heart rate series. Biomed Phys Eng Express. 2023;9.
30. Dimitriev DA, Saperova EV, Dimitriev AD. State anxiety and nonlinear dynamics of heart rate variability in students. PLoS One. 2016;11:e0146131.
31. Corbett M. From law to folklore: work stress and the Yerkes-Dodson Law. J Manag Psychol. 2015;30:741-52.
32. Kim HJ, Park Y, Lee J. The validity of heart rate variability (HRV) in educational research and a synthesis of recommendations. Educ Psychol Rev. 2024;36:42.
33. Marazziti D, Di Muro A, Castrogiovanni P. Psychological stress and body temperature changes in humans. Physiol Behav. 1992;52:393-5.
34. Colbert AP, Spaulding K, Larsen A, Ahn AC, Cutro JA. Electrodermal activity at acupoints: literature review and recommendations for reporting clinical trials. J Acupunct Meridian Stud. 2011;4:5-13.
35. Li S, Sung B, Lin Y, Mitas O. Electrodermal activity measure: a methodological review. Ann. Tour. Res. 2022;96:103460.
36. Benedek M, Kaernbach C. A continuous measure of phasic electrodermal activity. J Neurosci Methods. 2010;190:80-91.
37. Tronstad C, Amini M, Bach DR, Martinsen ØG. Current trends and opportunities in the methodology of electrodermal activity measurement. Physiol Meas. 2022;43.
38. Cecchi S, Piersanti A, Poli A, Spinsante S. Physical stimuli and emotions: EDA features analysis from a wrist-worn measurement sensor. IEEE. 2020.
39. Lee J, Hwang HB, Lee S, et al. Analysis of acute stress reactivity and recovery in autonomic nervous system considering individual characteristics of stress using HRV and EDA. IEEE Access. 2024.
40. Klimek A, Mannheim I, Schouten G, Wouters EJM, Peeters MWH. Wearables measuring electrodermal activity to assess perceived stress in care: a scoping review. Acta Neuropsychiatr. 2023;37:e19.
41. K Kasap A, Kurt B. Exploring the correlation of physiological stress signals with student exam performance: a preliminary study. Appl Psychophysiol Biofeedback. 2025;50:149-64.

DOI: 10.4274/anatoljmed.2026.51523
Anatol J Gen Med Res 2026;36(2):142-51

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Correlation Between Perfusion Index, Oxygen Saturation and Cord Blood Gas in Infants Delivered Vaginally or with Cesarean Section

Vajinal veya Sezaryen Doğumla Doğan Yenidoğanlarda Perfüzyon İndeksi, Oksijen Satürasyonu ve Kordon Kan Gazı Arasındaki İlişki

Merve Yavuz[1], Esin Koç[2], İbrahim Hirfanoğlu[2], Ebru Ergenekon[2], Eray Esra Önal[2], Canan Türkyılmaz[2]

[1]Gazi University Faculty of Medicine, Department of Pediatrics, Ankara, Türkiye
[2]Gazi University Faculty of Medicine, Department of Pediatrics, Division of Neonatology, Ankara, Türkiye

Cite as: Yavuz M, Koç E, Hirfanoğlu İ, Ergenekon E, Önal EE, Türkyılmaz C. Correlation between perfusion index, oxygen saturation and cord blood gas in infants delivered vaginally or with cesarean section. Anatol J Gen Med Res. 2026;36(2):142-51

Abstract

Objective: Determining fetal/neonatal well-being by monitoring physiological changes, evaluating oxygenation/alveolar ventilation and circulation/perfusion, and the need for resuscitation are considered necessary during the fetal/neonatal transition period after delivery. This study aims to investigate the correlation between non-invasive methods oxygen saturation (SpO_2), perfusion index (PI), and cord blood gas parameters during the postnatal transition in term or near-term infants delivered vaginally or with cesarean section.

Methods: Sixty-four healthy term/near term infants were enrolled into this study. The blood gases of the umbilical artery (UA) and umbilical vein (UV) were analyzed in each case. SpO_2 and PI were recorded with a pulse oximeter.

Results: SpO_2 and PI values of vaginal births were significantly higher than who were born with cesarean section. A significant correlation between PI and SpO_2 values of both UA and UV was found in the vaginal birth group, while no correlation was found in the cesarean group. The trend of the PI was time-independent, and PI reached stability before SpO_2.

Conclusion: Non-invasive PI and SpO_2 measurements could be promising in monitoring the oxygenation and circulation of neonates during the postnatal transition period.

Keywords: Neonates, newborn, perfusion index, transition period, umbilical cord blood gases

Öz

Amaç: Doğum sonrasında gerçekleşen fetal/neonatal geçiş döneminde fizyolojik değişikliklerin izlenmesi, oksijenasyon/alveolar ventilasyon ve dolaşım/perfüzyonun değerlendirilmesi ile resüsitasyon gereksiniminin belirlenmesi, fetal/yeni doğan iyilik halinin saptanmasında gerekli kabul edilmektedir. Bu çalışma, vajinal ya da sezaryen ile doğan miadında veya miada yakın bebeklerde, postnatal geçiş sürecinde non-invaziv yöntemlerle ölçülen oksijen satürasyonu (SpO_2), perfüzyon indeksi (PI) ve kordon kan gazı parametreleri arasındaki korelasyonu araştırmayı amaçlamaktadır.

Address for Correspondence/Yazışma Adresi: Merve Yavuz, MD, Gazi University Faculty of Medicine, Department of Pediatrics, Ankara, Türkiye
E-mail: drmerveyavuz@hotmail.com
ORCID ID: orcid.org/0000-0002-7143-0943

Received/Geliş tarihi: 21.11.2025
Accepted/Kabul tarihi: 20.02.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

Yöntem: Bu çalışmaya 64 sağlıklı miadında/miada yakın yenidoğan dahil edilmiştir. Her bir olguda umbilikal arter (UA) ve umbilikal ven (UV) kan gazları analiz edilmiştir. SpO_2 ve PI değerleri pulse oksimetre ile kaydedilmiştir.

Bulgular: Vajinal doğumla dünyaya gelen bebeklerin SpO_2 ve PI değerleri, sezaryen ile doğanlara kıyasla anlamlı derecede daha yüksek bulunmuştur. Vajinal doğum grubunda PI ve SpO_2 değerlerinin hem UA hem UV kan gazı parametreleriyle anlamlı korelasyon gösterdiği, buna karşın sezaryen grubunda herhangi bir korelasyon saptanmadığı belirlenmiştir. PI'nın zamandan bağımsız bir eğilim gösterdiği ve PI'nin SpO_2 değerlerinden daha önce stabil düzeye ulaştığı görülmüştür.

Sonuç: Non-invaziv PI ve SpO_2 ölçümleri, yenidoğanların postnatal geçiş döneminde oksijenasyon ve dolaşımın izlenmesinde umut vaat eden yöntemler olabilir.

Anahtar Kelimeler: Neonatoloji, yenidoğan, perfüzyon indeksi, geçiş dönemi, umbilikal kordon kan gazları

Introduction

The early postnatal transition a period during which significant changes occur as circulation shifts from fetal to neonatal circulation. During this time, when the umbilical cord is cut, the placental blood flow stops, the lungs fill with air, and gas exchange begins, more oxygen (O_2) is carried to the tissues, the systemic vascular resistance increases, and the intracardiac shunts (foramen ovale, ductus arteriosus) are closed. This adaptation process is primarily influenced by the gestational age, the mode of delivery, and the type of maternal anesthesia[1]. Therefore, during this period after birth, it is crucial to obtain information about fetal/neonatal well-being, to monitor physiological changes during the transition, and to evaluate oxygenation/alveolar ventilation, circulation, and the need for resuscitation. Hence, we used umbilical cord blood gas analysis and some non-invasive methods for these purposes.

Hemoglobin oxygen saturation (SpO_2), and perfusion index (PI) measurements with a pulse oximeter are recommended in many guidelines, such as newborn care, resuscitation, and screening of congenital heart diseases (CHDs). They are widely used in current clinical practice during and after delivery, including in neonatal intensive care units. SpO_2 and PI measurement are preferred because they are non-invasive especially in monitoring these transitional changes in the early postnatal period and determining the need for intensive care[2-7]. The PI is an important non-invasive indicator of peripheral circulation in clinical practice for newborns. Hemodynamic and respiratory changes during the transition period may affect peripheral perfusion[8]. These findings suggest that the PI is not affected by the mode of delivery in healthy term newborns[9]. Obtaining and analyzing umbilical cord blood gases immediately after delivery is the most important method for evaluating fetal and neonatal well-being[10-12]. There are a few study investigating the effect of mode of delivery and anaesthesia on the non-invasive monitorization parameters (SpO_2/PI) and umbilical artery (UA)/umbilical vein (UV) blood gas values in the transition period of healthy term infants.

This study aimed to investigate the relationship and compatibility between umbilical cord blood gas values and non-invasive measures (such as SpO_2 and PI), as well as delivery and anesthesia methods, during the transition period in term and near-term infants.

Materials and Methods

Sixty-four healthy term newborn infants (38 girls, 26 boys) born by [vaginal route (VR); 20] or [cesarean section (C/S); 44], without any prenatal or natal risk factors, were included in the study between February and April 2017. In our country, in non-emergent situations where urgent C/S is not necessary, the decision regarding the mode of delivery is left to the mother's preference. Our study included only infants born to women who underwent elective C/S. Individuals requiring resuscitation at birth and those with Apgar scores below 8 at the 5th minute were excluded from the study. Ethical approval was obtained from the Ethics Committee of Gazi University Faculty of Medicine (approval number: 2017-52, date: 19.01.2017). Written informed consent was obtained from the parents of all newborns included in the study.

Demographic characteristics of all infants included in the study and maternal characteristics were recorded. The gender, gestational week, birth weight, height, and head circumference of newborns; type of delivery; type of anesthesia applied at birth; maternal age; first- and fifth-minute Apgar scores; and postpartum interventions were recorded. All newborns were examined by a pediatrician in the operating or delivery room according to the mode of delivery.

After birth, 0.5-1 mL of blood was taken from the UA and the UV of each newborn infant using a special dry heparin blood gas injector, which was clamped on both sides. "Radiometer ABL 800 Basic blood gas analyzer" working with an ion-selective electrode system was used to determine blood gas parameters (pH, SpO_2 and partial pressure of carbon dioxide (pCO_2) immediately. Blood gases were evaluated.

After stabilization and evaluation of Apgar scores, Masimo set pulse oximeter rad-5v probe was attached to the right arm of the infants (preductal), SpO_2 and PI values were recorded continuously for the first 20 minutes.

Statistical Analysis

Mean, standard deviation, median, minimum, maximum, and percentile values in descriptive statistics for continuous data; percent values are given in the discrete data.

Pearson and Spearman correlation coefficients were used to evaluate the relationship of the UA and UV measurements with SpO_2, PI measurements at the 5th, 10th, 15th, and 20th minutes.

Either the t-test or the Mann-Whitney U test was used to compare data obtained from measurements of patients born by C/S and VR, after evaluating the conformity of the data to the normal distribution.

The Kaplan-Meier curve was used to calculate the stabilization time of the PI and SpO_2 measurements. P<0.05 was considered statistically significant.

Results

The mean gestational age of 64 cases was 38.6±1.1 (37-41 weeks). The delivery modes of the cases were C/S in 44 (69%) and vaginal in 20 (31%). General anesthesia was administered four (9%) C/Ss, and spinal anesthesia was administered in epidural anesthesia was used in four (20%) vaginal births, while induction was performed in nine (45%) vaginal births. The fifth-minute Apgar score was ≥9 in 93% of the infants. There have been no cases in which postpartum resuscitation was required.

Comparison of cord blood gas values by delivery type is shown in Table 1. Among those whose birth type was C/S or VR, there were significant differences between UA and UV in pH and pCO_2 values (Table 1). Among those whose birth type was C/S, both UA and UV pH values were significantly higher than in those with VR (p<0.001) (Table 1). In those with VR delivery, both UA and UV pCO_2 values were significantly higher than in those with C/S (p<0.05). There were no differences in UA and UV SpO_2 values between those with C/S and VR delivery types (p>0.05) (Table 1).

The median of preductal SpO_2 values at 10, 15, and 20 minutes in VR was significantly higher than in those born by C/S (Table 2). There was no difference in 5th-minute SpO_2 values between those delivered by C/S and those delivered by VR (p>0.05) (Table 2).

The comparison of the 5th-, 10th-, 15th-, and 20th-minute preductal PI averages , measured by the Masimo device, according to the mode of delivery, is shown in Table 3. There was no difference in terms of 5th and 10th-minute PI values in those whose mode of delivery was C/S or VR (p>0.05) . The difference between the 15th- and 20th-minute PI values was significant in those whose delivery type was C/S or VR (p<0.05). The median PI values at the 15th and 20th minutes were significantly higher in those with VR delivery (Table 3).

A positive correlation was observed between the 5th-min UA pO_2 value and the 5th-, 10th-, 15th-, and 20th-min PI values in infants born vaginally (p<0.05). There was no significant correlation between UA pO_2 and PI values at 10, 15, and 20 minutes (p>0.05) (Table 4). A positive correlation was also found between UV pO_2 and 10th- and 15th-minute PI values in vaginally born infants (p<0.05) (Table 4). There was no relationship between cord blood gas parameters and PI in those born by C/S.

Table 1. Comparison of cord blood gas parameters of cases according to delivery type

		C/S (n=44)		Vaginal route (n=20)		
		Mean ± SD (min-max)	Median	Mean ± SD (min-max)	Median	p-value*
Umbilical artery	pH	7.32±0.04 (7.18-7.39)	7.32	7.27±0.04 (7.2-7.37)	7.28	**0.000**
	pCO_2	46.1±5.8 (37.3-59.6)	45.2	49.6±5.3 (37.8-58.6)	50	**0.029**
	SpO_2	22.3±4.2 (13.4-29.9)	23	24.4±9.8 (14.1-46)	19.5	0.249
Umbilical vein	pH	7.36±0.04 (7.22-7.42)	7.36	7.32±0.04 (7.24-7.40)	7.32	**0.000**
	pCO_2	40.5±5.2 (31.7-54.6)	39.1	43.6±5.0 (33-52)	44.4	**0.029**
	SpO_2	30.9±5.2 (23.4-45.8)	30	31.5±8.7 (20-49)	29.4	0.727

*: T-test, pCO_2: Pressure of carbon dioxide, SpO_2: Oxygen saturation, SD: Standart deviation, C/S: Cesarean section

Table 2. Comparisons of preductal SpO_2 measurements by delivery type

	C/S	VR	
SpO_2	**Median (min-max)**	**Median (min-max)**	**p-value***
5th min	96 (83-100)	96 (84-100)	0.610
10th min	97 (80-100)	98.5 (86-100)	**0.025**
15th min	97 (86-100)	99 (84-100)	**0.027**
20th min	97.5 (89-100)	100 (88-100)	**0.020**

*: Mann-Whitney U test, SpO_2: Oxygen saturation, C/S: Cesarean section, VR: Vaginal route

Table 3. Comparison of perfusion index measurements according to delivery type

	C/S		VR		
PI	**Mean ± SD**	**Median (min-max)**	**Mean ± SD**	**Median (min-max)**	**p-value***
5th min	2.1±0.7	1.9 (0.5-3)	2.2±0.7	2 (1.25-3)	0.539
10th min	2.1±0.9	1.9 (1-5)	2.4±1.1	2 (1.25-5)	0.361
15th min	1.9±0.6	1.8 (1.25-3)	2.4±0.9	2 (1.25-5)	**0.009**
20th min	1.9±0.7	1.5 (0.50-3)	2.2±0.6	2 (1.25-3)	**0.012**

*: Mann-Whitney U test, C/S: Cesarean section, VR: Vaginal route, SD: Standard deviation

Table 4. The relationship between umbilical artery (UA) and vein (UV) pH, PCO_2 and SpO_2 values and perfusion index values in babies born by vaginal route

PI		UA pH	UA pCO_2	UV pH	UV pCO_2	UV SpO_2	UA SpO_2
5th min	r*	0.05	-0.04	0.23	-0.23	0.43	0.49
	p	0.85	0.88	0.32	0.33	0.06	**0.029**
10th min	r*	0.2	-0.18	0.11	0.13	0.57	0.4
	p	0.4	0.44	0.65	0.57	**0.009**	0.08
15th min	r*	0.22	0.03	0.43	-0.09	0.49	0.24
	p	0.36	0.9	0.06	0.69	**0.026**	0.31
20th min	r*	0.09	0.13	0.16	0.05	0.26	0.08
	p	0.72	0.59	0.52	0.83	0.27	0.75

*: Spearman correlation coefficient, pCO_2: Pressure of carbon dioxide, SpO_2: Oxygen saturation, PI: Perfusion index

Time-dependent univariate linear regression model results of the first quarter, second quarter, and third quarter values for pulse, SpO_2, and PI in the first 10 minutes, showed that time had no role in the change of PI, while the effect of time was high in explaining the change of pulse and SpO_2 (Table 5) (Figure 1). The PI and the stability times for SpO_2 were calculated using the Kaplan-Meier curve. The stability time within 5 consecutive sections for PI (less than 20% change) and SpO_2 (less than 1% change) was measured. The PI reached a stable state by one minute postnatal age in 37 newborns (59.7%). SpO_2 stabilized one minute after birth in 15 newborns (25.4%). While the PI reached a stable state within ≤5 minutes in 88.7% of cases, SpO_2 reached a steady state within ≤5 minutes in 52.6% of cases (Figure 2).

Discussion

The relationships between umbilical cord venous and umbilical cord arterial blood gas parameters at birth and PI and SpO_2 values measured by pulse oximetry in vaginal and C/S deliveries were investigated during postnatal transition in the present study. Our results showed that, SpO_2 and PI values of vaginal births were significantly higher than those, born with C/S. A significant correlation between PI and SpO_2 values in both UA and UV was found in the vaginal group, while no correlation was found in the caesarean group. The trend of PI was time independent, and PI reached stability before SpO_2. The evaluation of the compatibility between these measurements allowed discussion of how much it

Table 5. Time-dependent univariate linear regression model results of 25th, 50th, and 75th percentile values for perfusion index, SpO_2, and heart rate in the first 10 minutes

			R^2
SpO_2	1st quartile (25.p*)	Y=90.17+0.61*x	0.679
	2nd quartile (50.p*)	Y=93.97+0.37*x	0.778
	3rd quartile (75.p*)	Y=96.47+0.28*x	0.802
Pulse	1st quartile (25.p*)	Y=144.27-0.44*x	0.591
	2nd quartile (50.p*)	Y=153.97-0.74*x	0.895
	3rd quartile (75.p*)	Y=166.03-1.10*x	0.853
PI	1st quartile (25.p*)	Y=1.52-0.002*x	0.030
	2nd quartile (50.p*)	Y=2.00-0.002*x	0.030
	3rd quartile (75.p*)	Y=2.983-0.002*x	0.003

*: Percentile, SpO_2: Oxygen saturation, PI: Perfusion index

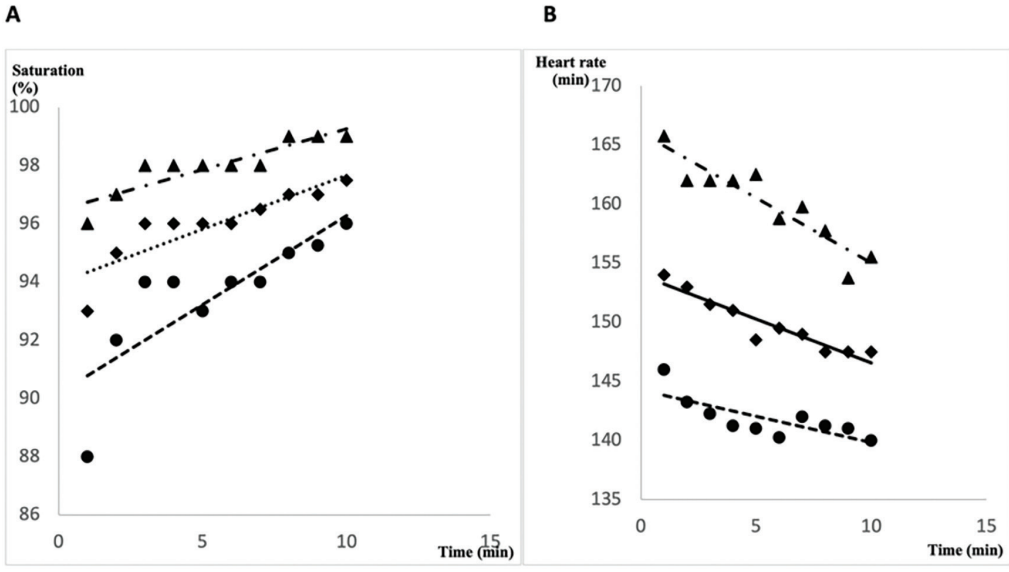


Figure 1. Time-dependent graph of 1st, 2nd and 3rd quartile values for oxygen saturation in the first 10 minutes (A); for heart rate in the first 10 minutes (B) ▲: 75.percentile, ◆: 50.percentile, ●: 25.percentile

could help in obtaining information about fetal well-being and early postnatal transition, and in determining the appropriate degree of respiratory support when resuscitation is required.

The transitional period in newborns is a critical interval during which significant changes occur in the transition from fetal circulation. During this period, a series of rapid physiological and anatomical changes occurs: placental blood flow ceases with clamping of the umbilical cord; the lungs begin gas exchange; pulmonary vascular resistance decreases; and intracardiac shunts close as systemic vascular resistance increases[1]. Few studies have examined the monitoring of infants during this period. Our study aimed to observe changes during the transition period and to evaluate the compatibility of pulse oximeter measurements with the blood gas data obtained from UA and UV by the invasive route.

Some methods have been developed that provide information on the well-being and prognosis of newborns and guidance for postpartum interventions. The most important of these is the analysis of umbilical cord blood gas parameters, which has become increasingly common. Apgar scoring, used in the

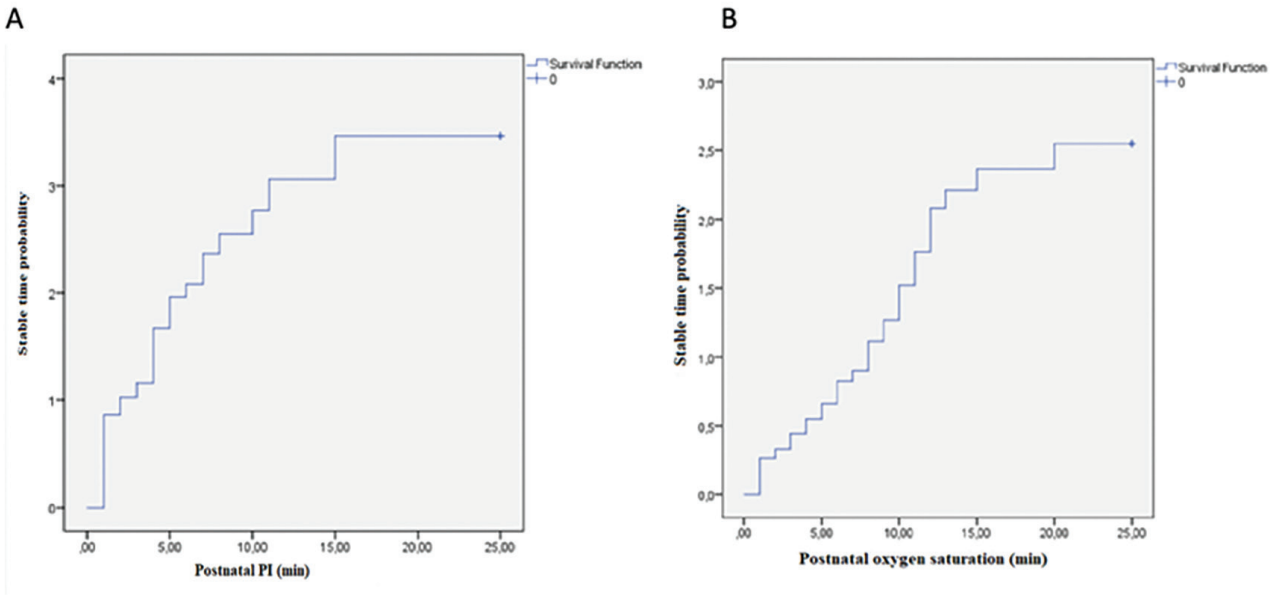


Figure 2. Kaplan-Meier curve for time to stable measurement for perfusion index (A), of oxygen saturation (B)

postnatal evaluation of newborns, helps distinguish normal newborns from those with noticeable signs of perinatal asphyxia. However, Apgar scoring is not sensitive enough to detect less severely affected infants. Clinical evaluation of newborns, combined with umbilical cord blood gas analysis after birth, guides early treatment approaches[13,14]. Arterial blood gas measurement is the gold-standard method for demonstrating acid-base disorders and for monitoring alveolar ventilation and oxygenation. The correction of acid-base and oxygenation disorders has a significant impact on mortality and morbidity[15]. Umbilical cord blood gas measurement is currently the most widely accepted method for evaluating fetal well-being and detecting perinatal asphyxia. However, in some emergency situations, umbilical cord blood gas samples cannot be collected or evaluated because of difficulties with blood withdrawal and technical problems, as collection is an invasive procedure. Technical issues, such as improper removal of blood gas, excess heparin in the injector, residual air in the blood sample, or inability to operate immediately, can cause incorrect results. Moreover, cord blood obtained from the UV or from mixed samples, rather than from the UA, may reflect the mother's condition rather than the infant's. Hence, non-invasive, validated, quantitative markers or measurements are needed to evaluate hypoxia and perfusion in term infants during the early postnatal transition period.

Pulse oximetry measures the percentage of oxygenated hemoglobin using the difference in absorbance between red and infrared light and calculates the ratio of oxygenated hemoglobin in the pulsatile component. When tissue perfusion decreases, the skin, subcutaneous tissues, muscles, and gastrointestinal tract are initially affected, whereas vital tissues (brain, heart, adrenal glands) are protected temporarily. Thus, detection of decreased perfusion in nonvital tissues could be used as an early marker[16-19].

The PI, representing variable light absorption caused by pulsatile arterial blood flow reflects the ratio relative to the constant light absorption from non-pulsatile components, including venous blood, connective tissue, skin, bone, and other tissues. It reflects momentary changes in peripheral blood flow[19,20]. The PI is a promising non-invasive method for assessing peripheral capillary circulation and, indirectly, organ perfusion[21]. Soon, thanks to the development and widespread use of devices that measure advanced PI, quantitative monitoring of the transition period from birth will be possible. Since the 2000s, many studies have been conducted to evaluate the PI in assessing neonatal well-being and in the early detection of life-threatening conditions. Many studies have determined PI threshold values in both term and preterm infants[22]. PI levels below the threshold value in neonatal intensive care units have been proposed as objective indicators of acute disease. For instance, some

studies have shown that the PI value is low in CHDs, and its measurement has been suggested for diagnosing these diseases[20,23,24].

Schena et al.[23] screened 42.169 healthy term newborns by measuring PI and SpO_2 and detected CHD in three of them. Two of these were found to have low SpO_2, and one had a low PI. This study showed that PI and SpO_2 should be used in CHD screening[23,25-27]. Recent evidence supports adding PI to pulse oximetry to improve screening performance and reduce false positives in selected protocols[26,27]. Granelli and Ostman-Smith[20] measured preductal and postductal PI in 1000 healthy newborns within the first 1-120 hours of their lives and reported a mean PI of 1.70 (1.18-2.5). In that study, heart anomalies were detected in two infants with PI values below 0.7. In nine infants followed up as a control group for left heart obstruction disease, the PI was found to be below 1. Of them, five had PI values below 0.7. They stated that the PI indicates the severity of cardiac dysfunction and that a PI below 0.7 can be used as a marker for detecting CHDs[24].

Perrone et al.[28] also aimed to evaluate the role of peripheral PI and pulse oximetry screening during the first 24 hours of life in the early detection of CHDs and non-CHDs (perinatal infection and suspected respiratory disease) in newborns. This study concluded that combining peripheral saturation with the peripheral PI during the first 24 hours of life is predictive of minor CHDs and neonatal clinical conditions that require attention[28].

In our study, the PI was measured and recorded preductally at five-minute intervals for twenty minutes. Mean PI values were 2 (0.5-3) at the 5th minute, 2 (1-5) at the 10th minute, 1.75 (1.25-5) at the 15th minute, and 1.75 (0.5-3) at the 20th minute. The mean values were similar to those reported in the literature[9,23,24]. No apparent changes in the PI were observed within the first twenty minutes. De Felice et al.[24] also measured the PI between 0 and 1 minute and between 1 and 5 minutes after birth and reported no difference between those values. In our study, the lowest measured PI was 0.5. However, as it increased during follow-up, it was thought that this could be due to cold hands and feet resulting from vasomotor imbalance, which causes physiological peripheral cyanosis in newborns, along with vasospasm due to hypothermia. Nevertheless, hypothermia was excluded, as all measurements were made in incubators at ideal temperatures. In some measurements, PI exceeded 5; this was attributed to the babies being very active and crying during the measurement, because babies' crying can cause vasodilation and thus elevate PI.

Unal et al.[8] found the mean postductal PI at 10 minutes and at 1 hour to be 1.41 (1.17-1.78) in tachypneic and healthy term newborns. There was no significant difference in PI values between healthy and tachypneic infants[8]. According to our study, the lower measurements could be due to postductal measurement sites. On the other hand, in our study, we were not able either to to measure preductal and postductal PI simultaneously or to monitor body temperature continuously due to technical limitations. Hakan et al.[22] measured preductal and postductal SpO_2, PI, and pulse in the first five days of life in 196 terms and 45 preterm newborns. They reported the mean pre-ductal and postductal PI, measured over the first 6 hours of life in term infants, as 1.35 (1.02-1.91) and 0.88 (0.62-1.22), respectively. They stated that preductal PI was significantly higher than postductal PI in the first three days, and attributed the low values to regional and racial differences. As in our study, they observed fluctuations in PI during crying and activity[22].

In a study of 125 healthy term infants, Yiğit et al.[29] found that PI was higher in C/S births. This condition was assessed as indicative of more significant hemodynamic changes during the early transitional period after C/S delivery[29]. Our study found that 15th- and 20th-min PI values in those born vaginally were significantly higher than in those born by C/S. This could be attributed to vaginal delivery being a more physiological process. In our study, spinal anesthesia was applied in births with C/S where it adversely affects uteroplacental circulation and oxygenation. Although the duration of the effects of anesthetic agents on the neonate is unknown, a low PI in infants born by C/S may indicate circulatory disorders related to anesthesia.

In a study by Taguchi et al.[16], PI, pulse rate, and SpO_2 were measured for 10 minutes after birth in 60 newborns born between 35 and 41 weeks. The PI was 1.5 after birth in 45% of newborns, and in 90% of newborns it was measured as stable after 3 minutes. The stable PI value was always obtained before the SpO_2 measurement. It has been reported that there is no relationship between the course of PI and heart peak beat and SpO_2[30]. Our study also found that PI stabilized earlier than SpO_2. While the PI reached a stable state after the first minute postpartum in 37 newborns (59.7%), SpO_2 did so in only 15 newborns (25.4%). It was thought that PI would be low due to poor peripheral perfusion immediately after birth and would increase with time, such as SpO_2 or pulse. So we looked at how PI, SpO_2, and pulse changed over time in the first 10 minutes. At the tenth minute, 75% of the pulses were ≥155/min, 50% ≥145/min, and 25% ≤140/min

with 75% of PIs ≥3, 50% ≥2, 25≥1.5. It was seen that 75% of the SpO_2s were ≥98, 50% were ≥96 and 25% were 94. The measured values were consistent with the SpO_2 target table for the first 10 min, as determined by neonatal resuscitation program[6,7]. Additionally, recent percentile data among term infants show higher early SpO_2 levels under specific cord-clamping conditions[31]. If a PI target table can be created during the transition period, monitoring oxygenation and circulatory status and managing resuscitation steps can help inform decisions on intervention and hospitalization for the neonates. Contrary to the expectations from our study, PI followed a different course apart from the changes in pulse, and SpO_2 and remained in approximately the same value ranges after birth. A high PI indicates adequate perfusion of each organ immediately after birth. PI fluctuations observed from the first measurement may reflect individual differences in postnatal adaptability.

Changes in pulse and SpO_2 showed that the effect of time was significant, unlike in PI. Taguchi et al.[16] reported that infants born vaginally had higher SpO_2. While some studies indicate that SpO_2 is higher in vaginal delivery, some studies have shown no significant difference in SpO_2 according to the mode of delivery[32-36]. In our study, the SpO_2 values of the 10th, 15th, and 20th min vaginal births were significantly higher than those born with C/S. This situation can be explained by the more physiologic transition from fetal to neonatal circulation that occurs during vaginal delivery. In addition, the negative effects of anesthesia for C/S on circulation and oxygenation should not be ignored. However, we could not evaluate the effects of different forms of anesthesia because epidural and general anesthesia were insufficiently applied in the vaginal/CS groups.

Alderliesten et al.[36] followed up PI for 72 hours in 311 newborns younger than 32 weeks and found that PI was negatively related to dopamine, mean blood pressure, follow-up on mechanical ventilation, and SpO_2. However, it was positively associated with female gender, gestational age, pulse pressure, and patent ductus arteriosus (PDA). They suggested that the elevated pulse pressure in PDA caused blood to shunt into the pulmonary artery during diastole. Stroke volume is increased to eliminate the adverse effect of this condition on perfusion. PI was higher because peripheral vasodilation decreased diastolic blood pressure. A negative relationship between SpO_2 and PI was attributed to the effect of SpO_2 on vascular tone[36].

Our study found no relationship between mode of delivery (vaginal or C/S) and SpO_2 and PI values, in accordance with the study of Hakan et al.[22]. This situation might be due to the similarity between the two studies, both of which were conducted on healthy infants.

In the present study, we found a positive relationship between UA and UV SpO_2 and between UA and PI in vaginal births, while there was no relationship between umbilical cord blood gas parameters and PI in those delivered by C/S. To our knowledge, no studies in the literature have compared umbilical cord blood gas parameters with PI and SpO_2. The gold standard for demonstrating tissue oxygenation is arterial SpO_2. The PI is a parameter that helps us detect early deterioration of tissue perfusion, prevent tissue hypoxia and organ failure, and intervene in a timely manner. As a non-invasive method, the PI can evaluate blood flow and oxygenation simultaneously by measuring the ratio of pulsatile to non-pulsatile flow. Therefore, the positive relationship between SpO_2, the gold standard for demonstrating tissue oxygenation, and PI was considered an expected situation. The SpO_2 value, an indicator of oxygenation in both UV and UA blood gases, was positively associated with PI among infants born vaginally. It can be attributed to the fact that PI measurements reflect oxygenation more accurately during the transition from fetal to neonatal circulation, owing to the better physiological adaptation associated with vaginal delivery.

Study Limitations

This study has several limitations. First, the relatively small sample size and single-center design may limit the generalizability of our findings and reduce the statistical power to detect subtle differences between groups. Therefore, the results should be interpreted with caution. Second, the unequal distribution of participants between the vaginal and cesarean delivery groups and the limited number of cases receiving each type of anesthesia restricted our ability to perform detailed subgroup analyses of anesthesia effects. Third, preductal and postductal PI measurements could not be obtained simultaneously because of technical limitations, and continuous body temperature monitoring was unavailable during measurements; these limitations may have influenced peripheral perfusion values.

Future large-scale, multicenter studies with balanced group distributions and standardized anesthesia protocols are needed to confirm our findings and further clarify the relationship between PI, SpO_2, and umbilical cord blood gas parameters during the neonatal transition period.

Conclusion

This study evaluated the use of non-invasive pulse oximetry and PI in newborns during the postpartum transition period as a supplement to cord blood gas. In the near future, we hope it will be possible to use these non-invasive methods to support monitoring during the transition period and the resuscitation of newborns.

Ethics

Ethics Committee Approval: Ethical approval was obtained from the Ethics Committee of Gazi University Faculty of Medicine (approval number: 2017-52, date: 19.01.2017).

Informed Consent: Written informed consent was obtained from the parents of all newborns included in the study.

Footnotes

This study was derived from the medical specialty thesis of Merve Yavuz.

Authorship Contributions

Surgical and Medical Practices: M.Y., E.K., İ.H., E.E., E.Ö., C.T., Concept: E.K., C.T., Design: E.E., C.T., Data Collection or Processing: M.Y., Analysis or Interpretation: M.Y., İ.H., Literature Search: M.Y., Writing: M.Y.

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

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

References

1. Goldsmith JP. Overview and initial management of delivery room resuscitation. In: Martin RJ, Fanaroff AA, Walsh MC, editors. Fanaroff and Martin's neonatal-perinatal medicine: diseases of the fetus and infant. 10th ed. Philadelphia: Elsevier Saunders; 2015. p. 460-4.
2. Weiner GM, Zaichkin J, editors. Textbook of neonatal resuscitation. 7th ed. Elk Grove Village, IL: American Academy of Pediatrics and American Heart Association; 2016.
3. Clark JS, Votteri B, Ariagno RL, et al. Noninvasive assessment of blood gases. Am Rev Respir Dis. 1992;145:220-32.
4. Poets CF, Southall DP. Noninvasive monitoring of oxygenation in infants and children: practical considerations and areas of concern. Pediatrics. 1994;93:737-46.
5. Hess D. Detection and monitoring of hypoxemia and oxygen therapy. Respir Care. 2000;45:65-80; discussion 80-3.
6. Aziz K, Lee HC, Escobedo MB, et al. Part 5: neonatal resuscitation: 2020 American Heart Association Guidelines for Cardiopulmonary Resuscitation and Emergency Cardiovascular Care. Circulation. 2020;142:S524-50.
7. Madar J, Roehr CC, Ainsworth S, et al. European Resuscitation Council Guidelines 2021: newborn resuscitation and support of transition of infants at birth. Resuscitation. 2021;161:291-326.
8. Unal S, Ergenekon E, Aktas S, et al. Perfusion index assessment during transition period of newborns: an observational study. BMC Pediatr. 2016;16:164.
9. Kroese JK, van Vonderen JJ, Narayen IC, Walther FJ, Hooper S, te Pas AB. The perfusion index of healthy term infants during transition at birth. Eur J Pediatr. 2016;175:475-9.
10. Johnson JW, Riley W. Cord blood gas studies: a survey. Clin Obstet Gynecol. 1993;36:99-101.
11. Johnson JW, Richards DS, Wagaman RA. The case for routine umbilical blood acid-base studies at delivery. Am J Obstet Gynecol. 1990;162:621-5.
12. Leinonen E, Gissler M, Haataja L, et al. Umbilical artery pH and base excess at birth are poor predictors of neurodevelopmental morbidity in early childhood. Acta Paediatr. 2019;108:1801-10.
13. Yeomans ER, Hauth JC, Gilstrap LC 3rd, Strickland DM. Umbilical cord pH, PCO2, and bicarbonate following uncomplicated term vaginal deliveries. Am J Obstet Gynecol. 1985;151:798-800.
14. Dinleyici E, Tekin N. Sağlıklı term ve preterm yenidoğanlarda umblikal arterasit-baz durumu ve laktat düzeyleri ve doğum şeklinin ilişkisinin değerlendirilmesi. Perinatoloji Dergisi. 2005;13:191-7.
15. Barry JS, Deacon J, Hernandez C, Jones MD Jr. Acid-base homeostasis and oxygenation. In: Carter BS, Gardner SL, Hernandez JA, Hines ME, editors. Merenstein & Gardner's handbook of neonatal intensive care. 8th ed. Elsevier; 2016. p. 145-57.
16. Taguchi Y, Hosono S, Kayama K, et al. Target value of oxygen saturation during the first 10 min after birth. Pediatr Int. 2017;59:1064-8.
17. Tobin MJ. Respiratory monitoring. JAMA. 1990;264:244-51.
18. MacDonald PH, Dinda PK, Beck IT, Mercer CD. The use of oximetry in determining intestinal blood flow. Surg Gynecol Obstet. 1993;176:451-8.
19. Lima A, Bakker J. Noninvasive monitoring of peripheral perfusion. Intensive Care Med. 2005;31:1316-26.
20. Granelli Ad, Ostman-Smith I. Noninvasive peripheral perfusion index as a possible tool for screening for critical left heart obstruction. Acta Paediatr. 2007;96:1455-9.
21. Mandala VK, Mendu SB, Bollaboina SKY, Kotha R Sr. Role of perfusion index and pulse variability index in the assessment of neonatal hemodynamics: a systematic review. Cureus. 2023;15:e48058.
22. Hakan N, Dilli D, Zenciroglu A, Aydin M, Okumus N. Reference values of perfusion indices in hemodynamically stable newborns during the early neonatal period. Eur J Pediatr. 2014;173:597-602.
23. Schena F, Picciolli I, Agosti M, et al. Perfusion index and pulse oximetry screening for congenital heart defects. J Pediatr. 2017;183:74-9.e1.
24. De Felice C, Del Vecchio A, Criscuolo M, Lozupone A, Parrini S, Latini G. Early postnatal changes in the perfusion index in term newborns with subclinical chorioamnionitis. Arch Dis Child Fetal Neonatal Ed. 2005;90:F411-4.
25. Siefkes H, Kair L, Tancredi DJ, et al. Oxygen saturation and perfusion index-based enhanced critical congenital heart disease screening. Am J Perinatol. 2020;37:158-65.
26. Jiang SL, Zhan YJ, Yan P, Yue Y, Tang J. Pulse oximetry and perfusion index screening for congenital heart defects: a systematic review and meta-analysis. Am J Perinatol. 2023;40:1611-7.
27. Lannering K, Elfvin A, Mellander M. Low false-positive rate of perfusion index as a screening tool for neonatal aortic coarctation. Acta Paediatr. 2021;110:1788-94.

28. Perrone S, Giordano M, De Bernardo G, et al. Peripheral saturation and perfusion index on the first day of life play a role in early discharge of healthy term newborns. Int J Clin Pract. 2022;2022:2887312.
29. Yiğit Ş, Takcı Ş, Bozkaya D, Yurdakök M. Perfusion index and pleth variability index in the first hour of life according to mode of delivery. Turk J Pediatr. 2018;60:421-5.
30. Padilla-Sánchez C, Baixauli-Alacreu S, Cañada-Martínez AJ, Solaz-García Á, Alemany-Anchel MJ, Vento M. Delayed vs immediate cord clamping changes oxygen saturation and heart rate patterns in the first minutes after birth. J Pediatr. 2020;227:149-56.e1.
31. Dawson JA, Kamlin CO, Vento M, et al. Defining the reference range for oxygen saturation for infants after birth. Pediatrics. 2010;125:e1340-7.
32. Harris AP, Sendak MJ, Donham RT. Changes in arterial oxygen saturation immediately after birth in the human neonate. J Pediatr. 1986;109:117-9.
33. Rabi Y, Yee W, Chen SY, Singhal N. Oxygen saturation trends immediately after birth. J Pediatr. 2006;148:590-4.
34. Dimich I, Singh PP, Adell A, Hendler M, Sonnenklar N, Jhaveri M. Evaluation of oxygen saturation monitoring by pulse oximetry in neonates in the delivery system. Can J Anaesth. 1991;38:985-8.
35. Porter KB, Goldhamer R, Mankad A, Peevy K, Gaddy J, Spinnato JA. Evaluation of arterial oxygen saturation in pregnant patients and their newborns. Obstet Gynecol. 1988;71:354-7.
36. Alderliesten T, Lemmers PM, Baerts W, Groenendaal F, van Bel F. Perfusion index in preterm infants during the first 3 days of life: reference values and relation with clinical variables. Neonatology. 2015;107:258-65.

DOI: 10.4274/anatoljmed.2026.02486
Anatol J Gen Med Res 2026;36(2):152-60

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Determination of COVID-19 Knowledge-attitude-awareness Levels Among Pregnant Women in a Tertiary Care Hospital in Western Türkiye

Batı Türkiye'de Bir Üçüncü Basamak Sağlık Kurumunda Gebeler Arasında COVID-19 Bilgi-tutum-farkındalık Düzeylerinin Belirlenmesi

Özge Nemli[1], Serhat Sarıkaya[1], Salih Atakan Nemli[2]

[1]University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of Obstetrics and Gynecology, İzmir, Türkiye
[2]İzmir Katip Çelebi University Faculty of Medicine, Department of Infectious Diseases and Clinical Microbiology, İzmir, Türkiye

Cite as: Nemli Ö, Sarıkaya S, Nemli SA. Determination of COVID-19 knowledge-attitude-awareness levels among pregnant women in a tertiary care hospital in Western Türkiye. Anatol J Gen Med Res. 2026;36(2):152-60

Abstract

Objective: The effect of coronavirus disease-2019 (COVID-19) on pregnancy remains controversial. Currently, literature investigating levels of knowledge, behavior, and attitudes regarding COVID-19 among pregnant women is scarce. This study aims to determine the knowledge, attitudes, and awareness levels of pregnant women followed up at the obstetrics and gynecology clinic of a tertiary care hospital in Western Türkiye.

Methods: Pregnant women admitted to a tertiary care hospital in İzmir, a metropolitan city in Türkiye, between May 1 and June 30, 2020, were enrolled in the study. The researchers used a questionnaire to measure COVID-19 knowledge, attitudes, and awareness levels. Anxiety level was evaluated using the generalized anxiety disorder-7 scale. All analyses were performed using SPSS v.25 (IBM Corp., Armonk, NY) at a significance level of 0.05.

Results: The median COVID-19 knowledge score of the study group was 13.00 (12.00-15.00). Negative correlations were detected between the COVID-19 knowledge score and the number of pregnancies, live births, and children. Conversely, the knowledge score increased with higher education levels among pregnant women and their partners. Similarly, higher family income levels were associated with higher knowledge scores. Severe anxiety was detected in 109 (26.2%) pregnant women.

Conclusion: Pregnant women in the study had substantial COVID-19 knowledge levels. Knowledge level was positively correlated with positive behaviors, post-pandemic attitudes, higher education levels, and higher soci-economic status.

Keywords: Pregnancy, COVID-19, knowledge, anxiety

Öz

Amaç: Koronavirüs hastalığı-2019 (COVID-19)'un gebelik üzerindeki etkisi tartışmalıdır. Literatürde gebelerde bilgi, davranış ve tutum düzeylerini araştıran çok fazla çalışma bulunmamaktadır. Bu çalışmanın amacı, Türkiye'nin batısında bulunan bir üçüncü basamak hastanenin kadın hastalıkları ve doğum bölümünde izlenen gebelerin COVID-19 hakkında bilgi-tutum-farkındalık düzeylerini belirlemektir.

Address for Correspondence/Yazışma Adresi: Assoc. Prof., Salih Atakan Nemli, İzmir Katip Çelebi University Faculty of Medicine, Department of Infectious Diseases and Clinical Microbiology, İzmir, Türkiye
E-mail: atanemli@yahoo.com
ORCID ID: orcid.org/0000-0002-2720-9870

Received/Geliş tarihi: 30.12.2025
Accepted/Kabul tarihi: 10.03.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

Yöntem: 1 Mayıs-30 Haziran 2020 tarihleri arasında üçüncü basamak bir hastaneye başvuran gebeler çalışmaya alındı. Araştırmacılar tarafından hazırlanan bir anket, COVID-19 bilgi, tutum ve farkındalık düzeylerini ölçmek için kullanılmıştır. Anksiyete, yaygın anksiyete bozukluğu-7 ölçeği ile değerlendirildi. İstatiksel analizler SPSS v.25 (IBM Corp., Armonk, NY) paket programı kullanılarak yapıldı. Analizlerde p-değerinin <0,05 olması istatiksel olarak anlamlı kabul edildi.

Bulgular: Çalışma grubunun COVID-19 hakkındaki medyan bilgi düzeyi puanı 13.00 (12.00-15.00) idi. COVID-19 bilgi düzeyi puanı ile gebelik sayısı, canlı doğum sayısı ve çocuk sayısı arasında negatif korelasyonlar tespit edildi. Gebelerin ve eşlerinin eğitim düzeyi yükseldikçe COVID-19 bilgi düzeyi puanı da yükseldi. Aile gelir düzeyi arttıkça COVID-19 bilgi düzeyi puanı da arttı. Gebelerin 109'unda (%26,2) şiddetli anksiyete saptandı.

Sonuç: Çalışma grubunu oluşturan gebelerin COVID-19 bilgi düzeyleri oldukça yüksek bulunmuştur. Bilgi düzeyi ile olumlu davranışlar, pandemi sonrası tutumlar, yüksek eğitim düzeyi ve yüksek sosyo-ekonomik düzey parametreleri arasında pozitif yönde bir ilişki olduğu görüldü.

Anahtar Kelimeler: Gebelik, COVID-19, bilgi, kaygı

Introduction

Coronaviruses (CoVs) are a large family of viruses that cause a variety of diseases, ranging from the common cold to more serious conditions such as middle east respiratory syndrome (MERS-CoV) and severe acute respiratory syndrome (SARS-CoV)[1]. CoVs cause zoonotic infections; however, many CoVs detected in animals have not yet been transmitted to humans. The subtypes of coronavirus currently circulating in humans are HCoV-229E, HCoV-OC43, HCoV-NL63, and HKU1-CoV[2].

On December 31, 2019, the World Health Organization (WHO) reported pneumonia cases of unknown etiology in Wuhan, China. Through sequence analysis of samples, a previously unknown novel coronavirus was discovered and named 2019-nCoV[3]. Soon after, this disease was defined as coronavirus disease-2019 (COVID-19) by the WHO[4]. Due to its structural similarity to SARS-CoV, the novel coronavirus was named SARS-CoV-2[5]. The first COVID-19 cases in Türkiye were reported on March 19, 2020[6].

The spectrum of infection ranges widely from mild to critical; however, most symptomatic infections are not severe[7]. COVID-19 infection is characterized by a wide spectrum of clinical manifestations, ranging from completely asymptomatic cases to severe acute respiratory distress syndrome (ARDS). Approximately 5-10% of patients with COVID-19 develop severe ARDS[8]. Severe infection and increased mortality rates are associated with older age or underlying comorbidities such as cardiovascular diseases, diabetes mellitus, hypertension, chronic lung disease, and cancer[9]. Currently, no specific treatment for COVID-19 has been proven effective. Except for remdesivir, no antiviral drug has been fully approved, although many agents have been investigated *in vitro* and in some *in vivo* studies, showing promising results[10].

It is known that physiological changes in the immune system during pregnancy can affect susceptibility to and the severity of infectious diseases[11]. In SARS and MERS-CoV experiences, pregnant women were observed to be a high-risk group for complications of coronavirus infections[12]. However, data regarding the effect of COVID-19 on pregnancy remain limited[13-15]. Most pregnant patients exhibit mild or moderate symptoms similar to non-pregnant women, and morbidity and mortality rates comparable to the general population have been observed[16,17].

This study aims to determine the levels of knowledge, attitudes, and awareness regarding COVID-19 among pregnant women attending the obstetrics and gynecology clinic of a tertiary care hospital in Western Türkiye.

Materials and Methods

This cross-sectional study was conducted between May 1 and June 30, 2020. The survey was conducted in the clinic of obstetrics and gynecology at a tertiary care hospital in Western Türkiye. The target population, consisting of pregnant women who sought pregnancy follow-up or were hospitalized in wards during the study period, was estimated to be 600 individuals. An a priori power analysis was conducted, and the minimum sample size was calculated to be 316, assuming a type 1 error of 1% (99% confidence) and a confidence interval of 5%. The final study group consisted of 416 pregnant women. A total of 21 pregnant women who declined to participate were excluded.

The questionnaire was administered by the researchers via face-to-face interviews. The first part of the questionnaire contained 13 questions regarding demographic characteristics, pregnancy, and birth history. In the second part, the question "Have you heard of the epidemic of the new coronavirus (COVID-19) disease?" was asked to assess

awareness. In the third, fourth, and fifth sections, knowledge levels regarding COVID-19 symptoms (5 questions, 3-point Likert scale), transmission pathways (4 questions, 3-point Likert scale), and treatment/prevention (9 questions, 2-point Likert scale, true/false) were evaluated. One point was awarded for each correct answer. The total COVID-19 knowledge score ranged from 0 to 18 points; higher scores indicated a higher level of knowledge.

In the sixth section, positive behaviors related to COVID-19 were evaluated using a 5-point Likert scale by asking about seven specific behaviors. The frequency of participants responding "I apply very well/well" was recorded, and correlations between this frequency and the COVID-19 knowledge score were analyzed. The seventh chapter examined participants' self-perception of the sufficiency of their knowledge and their sources of information about COVID-19.

In chapter eight, changes in attitudes following the onset of the pandemic were evaluated using 16 questions. The first 4 questions related to increased anxiety, while the other 5 related to self-restraint in social life. For these questions, the frequency of participants responding "I strongly agree/I agree" was presented, along with correlations to the knowledge score. Additionally, information regarding childcare and the stocking of materials after the start of the pandemic was presented as frequencies.

The anxiety levels of participants were evaluated using the generalized anxiety disorder-7 (GAD-7) scale[18], which has been adapted to Turkish with proven validity and reliability[19]. Total scores were categorized as follows: 0 points indicated "no anxiety"; 1-4 points indicated "mild anxiety"; 5-9 points indicated "moderate anxiety"; and 15 points and above indicated "severe anxiety".

The study group was divided into three categories based on gestational age: 1st trimester (1-13 weeks), 2nd trimester (14-26 weeks), and 3rd trimester (27-40 weeks). Marital status was categorized as married, cohabiting, or single/divorced/separated. Professions were grouped into the categories: housewife, unemployed, blue collar, and white collar. Spousal profession was categorized similarly. Monthly family income was categorized based on multiples of the minimum wage: below the minimum wage, around the minimum wage, 2-3 times the minimum wage, and 4 times the minimum wage or above.

This study was approved by University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Local Ethics Committee (approval no: 2020/6-20, date: 13.05.2020). All procedures were performed in accordance with the ethical standards of the institutional and/or national research committee. All participants provided written informed consent. Confidentiality was maintained throughout the study.

Statistical Analysis

IBM SPSS® (Statistical Package for the Social Sciences) version 25.0 (IBM Corp., Armonk, NY) was used for statistical analyses. The Normal distribution was evaluated using Q-Q plots, histograms, skewness and kurtosis values, and the Shapiro-Wilk test. As the data did not conform to a normal distribution, non-parametric tests were used. Frequency values were presented as n (%), and average values as median (25^{th}-75^{th} percentile). Spearman correlation analysis was used to assess the relationships between variables. A p-value <0.05 was considered statistically significant.

Results

The median age of the 416 pregnant women in the study group was 28. Those in the third trimester of pregnancy accounted for 210 (50.5%) of the participants. Primigravida women (first pregnancy) constituted 108 (25.9%) of the group. The proportion of women with one live birth was 38.2% (n=159). The median (25-75%) values for the number of pregnancies, live births, and children were 2 (1-3), 1 (0-2), and 1 (0-2), respectively.

In our study, 338 (93.3%) of the pregnant women were married, and 205 (49.3%) were primary- or secondary-school graduates. A large majority (n=398, 95.7%) lived in urban areas. Among the spouses of participants, 218 (52.4%) were primary or secondary school graduates, and 291 (70.0%) were blue collar workers. It was reported that 29.8% (n=124) of pregnant women had monthly incomes below the minimum wage. Socio-demographic characteristics of the study group are presented in Table 1.

All participants were aware of the COVID-19 pandemic. The three least-known propositions about the pandemic were: "People with the disease certainly have a complaint/finding" (27.2% correct); "All COVID-19 patients have serious symptoms" (31.5% correct); and "In order not to get COVID-19, we must try to stay at least 10 meters away from other people" (4.4% correct). Conversely, the most commonly known propositions were: "Touching the eyes, nose or mouth with dirty hands is risky" (90.9%); "Hands should be cleaned frequently with water-soap or alcohol-based disinfectants"

Table 1. The distribution of the study group according to some socio-demographic characteristics		
Socio-demographic characteristics	**Median (25-75%)**	**n (%)**
Age group (year)	28.00 (24.00-33.00)	
Trimester		
1st trimester		41 (9.9)
2nd trimester		165 (39.7)
3rd trimester		210 (50.5)
Number of pregnancy	2.00 (1.00-3.00)	
Number of live births	1.00 (0.00-2.00)	
Number of children	1.00 (0.00-2.00)	
Marital status		
Separate, single, divorced		5 (1.2)
Married		388 (93.3)
Lives with her partner		23 (5.5)
Educational level of the pregnant		
Illiterate		23 (5.5)
Primary or secondary school		205 (49.3)
High school		115 (27.6)
Higher education		73 (17.5)
Profession of the pregnant		
Housewife		313 (75.2)
Unemployed		12 (2.9)
Blue collar		46 (11.1)
White collar		45 (10.8)
Residential area		
Rural		18 (4.3)
Urban		398 (95.7)
Educational level of the husband/partner		
Illiterate		19 (4.6)
Primary or secondary school		218 (52.4)
High school		104 (25.0)
Higher education		75 (18.0)
Profession of the husband/partner		
Unemployed		42 (10.0)
Blue collar		291 (70.0)
White collar		83 (20.0)
Monthly family income		
Below the minimum wage		124 (29.8)
Around the minimum wage		205 (49.3)
Twice of the minimum wage		55 (13.2)
Three times the minimum wage		16 (3.8)
Four times the minimum wage and above		16 (3.8)

(91.1%); "When coughing/sneezing, mouth/nose should be covered with tissue or elbow" (91.1%); "One should try to stay at home except in compulsory situations" (91.1%); and "As far as possible, we should avoid touching the eyes, nose and mouth with our hands" (91.3%). Questions and affirmative answers are presented in Table 2.

The median COVID-19 knowledge score of the study group was 13(12-15), with minimum and maximum values of 0 and 18, respectively. No correlation was found between the COVID-19 knowledge score and age (r=0.064; p=0.195). However, negative correlations were observed between the knowledge score and the numbers of pregnancies, live births, and children (r=-0.152, p=0.002; r=-0.162, p=0.001; and r=-0.169, p=0.001, respectively).

No relationship was found between the knowledge score and either marital status or residential area (rural/urban). However, the knowledge score increased with the education level of pregnant women and their partners (r=0.415, p<0.001; r=0.284, p<0.001, respectively). Similarly, scores were higher when pregnant women and their partners worked in white collar jobs. As the family income level increased, the COVID-19 knowledge score increased (r=0.288; p<0.001). Correlations between knowledge scores and socio-demographic characteristics are presented in Table 3.

In the study, 77.4% (n=322) of participants reported their level of knowledge regarding the pandemic as sufficient. A positive correlation was found between the COVID-19 knowledge score and self-perceived knowledge sufficiency (r=0.209; p<0.001). Sources of information were television (93.8%), internet (54.1%), health workers (40.6%), social media (40.6%), mobile news applications (33.4%), and friends/relatives (31.3%).

The frequency of participants who reported performing the seven positive behaviors as "very well/well" ranged from 92.8% to 96.4%. Positive correlations were found between the application of each positive behavior rated "very well/well" and the COVID-19 knowledge score (p<0.001 for each). Frequency distributions for positive behaviors are presented in Table 4.

Regarding changes in attitudes, 53.0% of participants reported, "My concerns about my children increased. No correlation was found between anxiety regarding children and knowledge level (p>0.05). However, correlations were

Table. 2. Pregnant women's' knowledge regarding COVID-19

Questions	Affirmative answer, n (%)
The main symptoms of the disease are fever, weakness, dry cough and myalgia	362 (87)
Symptoms such as diarrhea, inability to taste and smell are not seen in this disease	179 (43)
Symptoms such as nasal congestion and runny nose are rarely seen	176 (42.3)
All COVID-19 patients have any complaint or finding	113 (27.2)
Most of COVID-19 patients have serious disease	131 (31.5)
It is transmitted by the inhalation of droplets of patients with coughing and sneezing	360 (86.5)
Virus can infect us, when we touch the contaminated places	370 (88.9)
Touching the eyes, nose or mouth with dirty hands is risky for the transmission of the disease	378 (90.9)
Although patients do not have a fever, they can infect other people	319 (76.7)
There is no effective cure for COVID-19 currently, but early symptomatic and supportive treatment may help patients to recover	340 (81.7)
An effective vaccine of the disease exists and is also applied in our country	219 (52.6)
To prevent the COVID-19 infection, people should avoid being in crowded places such as bazaars, buses, trains	372 (89.4)
No need to wear a mask in public places	333 (80)
Hands should be cleaned frequently with water-soap or alcohol-based disinfectants	378 (90.9)
As far as possible, we should avoid touching the eyes, nose and mouth with our hands	380 (91.3)
We must try to stay at least 10 meters away from other people	168 (40.4)
When coughing/sneezing, mouth/nose should be covered with tissue or elbow	379 (91.1)
It should be tried to stay at home except compulsory situations	379 (91.1)

COVID-19: Coronavirus disease-2019

found between other concerns and knowledge level (p<0.05). A positive correlation was found between self-restraint in social life (reported at >90% frequency) and knowledge level (p<0.05). The proportion of respondents stating "I try to stay away from the healthcare professionals I know" was 80.7%; no correlation was found between this avoidance and the knowledge score.

Among the 293 (70.4%) pregnant women with children, 70 (23.9%) reported needing additional childcare support during the pandemic. Of those not needing support, care was provided by themselves (17.9%), by older family members (20.6%), or by professional caregivers or friends (1.8%).

The median anxiety score (GAD-7) was 6.00 (2.00-10.00). Anxiety was absent in 68 (16.3%) women. Mild, moderate, and severe anxiety were found in 107 (25.7%), 132 (31.7%), and 109 (26.2%) women, respectively. No correlation was found between the anxiety score and the COVID-19 knowledge score (r=0.077; p=0.116).

Discussion

There is currently no evidence that pregnancy acts as a risk factor for COVID-19[14]. Literature investigating the levels of knowledge, behavior, and attitudes among pregnant women is limited. Our study offers an opportunity to assess these parameters among pregnant women in Türkiye.

Table 3. Correlations between COVID-19 knowledge level score and some socio-demographic characteristics

Spearman correlation (r; p) **Some socio-demographic characteristics**	**COVID-19 knowledge level score**
Age group (year)	r=0.064; p=0.195
Number of pregnancy	r=-0.152; **p=0.002**
Number of live births	r=-0.162; **p=0.001**
Number of children	r=-0.169; **p=0.001**
Marital status	r=-0.050; p=0.310
Residential area (rural/urban)	r=0.094; p=0.055
Educational level of the pregnant	r=0.415; **p<0.001**
Educational level of the husband/partner	r=0.284; **p<0.001**
Profession of the pregnant (Housewife, unemployed, blue collar,white collar)	r=0.143; **p=0.003**
Profession of the husband/partner (Unemployed, blue collar, white collar)	r=0.271; **p<0.001**
Monthly family income	r=0.288; **p<0.001**

COVID-19: Coronavirus disease-2019

Table 4. Frequency distributions for positive behaviors and their correlations with COVID-19 knowledge level score

Positive behaviors	**"Very good-good" practice, n (%)**	**Correlation with COVID-19 knowledge level score, (r; p)**
I try to stay away from crowded places (bazaar, market, bus, metro...) to prevent disease	386 (92.8)	r=0.349; **p<0.001**
I definitely wear a mask in public places	397 (95.5)	r=0.301; **p<0.001**
I often clean my hands with water-soap or alcohol-based disinfectants	391 (96.4)	r=0.238; **p<0.001**
As far as possible, I avoid touching the eyes, nose and mouth with my hand	396 (95.2)	r=0.191; **p<0.001**
I try to stay at least 1 meter away from other people	398 (95.6)	r=0.195; **p<0.001**
I cover my mouth/nose with handkerchief or elbow while coughing/sneezing	401 (96.4)	r=0.308; **p<0.001**
I try to stay at home, except when it is necessary	396 (95.2)	r=0.230; **p<0.001**

COVID-19: Coronavirus disease-2019

Table 5. Frequencies for attitudes which developed after the onset of the pandemic and their correlations with COVID-19 knowledge level scores

Attitudes which developed after the onset of the pandemic	I strongly agree-I agree, n (%)	Correlations with COVID-19 knowledge level scores, (r; p)
Attitudes (about concerns)		
My concerns about my own health/life have increased	360 (86.5)	r=0.263; **p<0.001**
I am worried about my unborn child	374 (89.9)	r=0.231; **p<0.001**
I am worried about my children	262 (53.0)	r=0.036; p=0.458
I have increased concerns about my elderly relatives (mom, dad... etc.)	374 (89.9)	r=0.290; **p<0.001**
Self-restraint in social life after the beginning pandemic		
I try to stay home as much as possible	397 (95.4)	r=0.283; **p<0.001**
I try not to use public transportation (bus, metro, etc.)	387 (93.0)	r=0.315; **p<0.001**
I try to stay away from social events such as visiting for relatives and friends	389 (93.5)	r=0.153; **p=0.002**
I try to stay away from all social activities	384 (92.3)	r=0.232; **p<0.001**
I try to stay away from the healthcare professionals who I know	336 (80.7)	r=0.045; p=0.363

COVID-19: Coronavirus disease-2019

All participants were aware of the COVID-19 pandemic. The average knowledge score was 13 out of 18, indicating a knowledge level of over 70%. In a study by Nwafor et al.[13] in Nigeria, the knowledge level regarding preventive measures was reported as good in 60.9% of pregnant women. Similarly, another study found that the majority of respondents had adequate knowledge about COVID-19 infection[20].

In our study, the level of knowledge decreased as the number of pregnancies, live births, and children increased. Although this could be affected by age, no correlation was found between age and knowledge level. Multiparity was also a factor reducing knowledge levels in the Nigerian study[13]. In Türkiye, multiparity is frequently observed in women with lower socio-economic status and lower education levels. This may be due to various cultural factors, as well as insufficient knowledge of, or access to, birth control methods.

Our results indicated that higher levels of COVID-19 knowledge were associated with higher education levels, white collar professions, and higher family income (Table 3), aligning with the results of the Nigerian study[13].

Mass media were the primary sources, whereas the proportion obtaining information from healthcare workers was relatively low (40%). Another study similarly reported that most participants received information from mass media, while only 20% received information from healthcare workers[20]. This suggests that healthcare professionals should increase educational efforts to provide adequate disease information.

The positive correlation between the study group's acceptance of their knowledge as "sufficient" and their actual knowledge scores suggests that participants consciously obtained information and followed pandemic developments. Knowledge level is a critical factor influencing attitudes and practices. Several studies have presented a positive correlation between knowledge levels and attitudes/practices regarding COVID-19[21,22]. Similarly, our study showed a positive correlation between knowledge level and the adoption of protective behaviors (Table 4). Conversely, the Nigerian study reported insufficiencies in the development of positive behaviors[13].

Anxiety, depression, and self-reported stress are common psychological reactions to the pandemic[23]. In this study, concerns about personal health, the unborn baby, and elderly relatives were high. Although vertical transmission and transmission via breast milk appear possible, current data have not yet definitively demonstrated COVID-19 positivity in fetal samples[24]. While correlations were observed between general anxiety and knowledge level, 53% of participants expressed concern for their children regardless of their knowledge level. Another study similarly found heightened

anxiety regarding older relatives, children, and unborn babies among pregnant women[25].

Approximately a quarter of pregnant women with children reported needing childcare support during the pandemic. This need should be addressed by governmental and non-governmental organizations, individually or jointly.

Our results showed that approximately 90% of pregnant women complied with social restrictions as suggested by the Turkish government and WHO[26,27], with compliance correlating with knowledge levels. Healthcare workers on the front lines may face stigma and discrimination[28], which may intensify as cases and mortality rates rise. Approximately 80% of our participants preferred to avoid contact with healthcare workers. Interestingly, as the level of knowledge increased, stigmatization decreased (Table 5). Disseminating precise information may help mitigate COVID-19-related stigma.

Managing health anxiety is crucial during pandemics[29]. Prenatal maternal distress can have negative impacts. Increased levels of anxiety and psychiatric symptoms have been reported in pregnant women during the pandemic[30]. Saccone et al.[31] found that more than half of pregnant women (53%) experienced severe psychological impacts. In our study, moderate-to-severe anxiety was observed in approximately 60% of participants. No correlation was found between anxiety levels and knowledge levels. However, Wang et al.[32] reported that while accurate health information was associated with lower stress levels, female gender and poor self-rated health were associated with higher psychological impact.

Study Limitations

Because of the rapid progression of the pandemic, a validity and reliability study of the questionnaire could not be conducted. Furthermore, this was a cross-sectional, single-center study. However, given the limited number of publications measuring knowledge, attitudes, and awareness among pregnant women, this study makes a significant contribution to the literature.

Conclusion

Little is known about the effects of COVID-19 on pregnancy. The COVID-19 knowledge levels of the pregnant women in this study were found to be substantial. Positive correlations were found between knowledge levels and positive behaviors, post-pandemic attitudes, education level, and socio-economic status. To protect the physical and psychological well-being of pregnant women and their babies, continuing education for health professionals is recommended. These results also highlight the need to focus on groups with lower knowledge levels.

Ethics

Ethics Committee Approval: This study was approved by University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Local Ethics Committee (approval no: 2020/6-20, date: 13.05.2020).

Informed Consent: All participants provided written informed consent.

Footnotes

Authorship Contributions

Concept: S.A.N., Design: S.A.N., Data Collection or Processing: Ö.N., S.S., Analysis or Interpretation: Ö.N., S.A.N., Literature Search: Ö.N., S.S., Writing: Ö.N., S.A.N.

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

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

References

1. Cascella M, Rajnik M, Aleem A, Dulebohn SC, Di Napoli R. Features, evaluation, and treatment of coronavirus (COVID-19). In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026.
2. Lau SK, Woo PC, Yip CC, et al. Coronavirus HKU1 and other coronavirus infections in Hong Kong. J Clin Microbiol. 2006;44:2063-71.
3. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med. 2020;382:727-33.
4. World Health Organization. Coronavirus disease (COVID-19) pandemic [Internet]. 2020. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019 [Accessed: December 30, 2025].
5. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol. 2020;5:536-44.
6. Republic of Turkey Ministry of Health. COVID-19 weekly situation report [Internet]. 2022. Available from: https://covid19.saglik.gov.tr
7. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet. 2020;395:497-506.
8. Rollas K, Şenoglu N. Covid-19 hastalarının yoğun bakım ünitesinde yönetimi. Anatol J Gen Med Res. 2020;30:142-55.
9. Petrilli CM, Jones SA, Yang J, et al. Factors associated with hospital admission and critical illness among 5279 people with coronavirus disease 2019 in New York City: prospective cohort study. BMJ. 2020;369:m1966.

10. Atalay S, Ersan G. Treatment of COVID-19. Tepecik Eğit ve Araşt Hast Dergisi. 2020;30:126-34.
11. Mor G, Cardenas I. The immune system in pregnancy: a unique complexity. Am J Reprod Immunol. 2010;63:425-33.
12. Li N, Han L, Peng M, et al. Maternal and neonatal outcomes of pregnant women with coronavirus disease 2019 (COVID-19) pneumonia: a case-control study. Clin Infect Dis. 2020;71:2035-41.
13. Nwafor JI, Aniukwu JK, Anozie BO, Ikeotuonye AC, Okedo-Alex IN. Pregnant women's knowledge and practice of preventive measures against COVID-19 in a low-resource African setting. Int J Gynaecol Obstet. 2020;150:121-3.
14. Rasmussen SA, Smulian JC, Lednicky JA, Wen TS, Jamieson DJ. Coronavirus disease 2019 (COVID-19) and pregnancy: what obstetricians need to know. Am J Obstet Gynecol. 2020;222:415-26.
15. Qiao J. What are the risks of COVID-19 infection in pregnant women? Lancet. 2020;395:760-2.
16. Public Health England. COVID-19: investigation and initial clinical management of possible cases [Internet]. 2020. Available from: https://www.gov.uk/government/publications/wuhan-novel-coronavirus-initial-investigation-of-possible-cases/investigation-and-initial-clinical-management-of-possible-cases-of-wuhan-novel-coronavirus-wn-cov-infection
17. Ellington S, Strid P, Tong VT, et al. Characteristics of women of reproductive age with laboratory-confirmed SARS-CoV-2 infection by pregnancy status — United States, January 22–June 7, 2020. Morbidity and Mortality Weekly Report. 2020;69:769-75.
18. Kessler RC, Keller MB, Wittchen HU. The epidemiology of generalized anxiety disorder. Psychiatr Clin North Am. 2001;24:19-39.
19. Konkan R, Şenormancı Ö, Güçlü O, Aydın E, Sungur MZ. Validity and reliability study for the Turkish adaptation of thegeneralized anxiety disorder-7 (GAD-7) scale. Archives of Neuropsychiatry. 2013;50:53-8.
20. Anikwe CC, Ogah CO, Anikwe IH, Okorochukwu BC, Ikeoha CC. Coronavirus disease 2019: knowledge, attitude, and practice of pregnant women in a tertiary hospital in Abakaliki, southeast Nigeria. Int J Gynaecol Obstet. 2020;151:197-202.
21. Zhong BL, Luo W, Li HM, et al. Knowledge, attitudes, and practices towards COVID-19 among Chinese residents during the rapid rise period of the COVID-19 outbreak: a quick online cross-sectional survey. Int J Biol Sci. 2020;16:1745-52.
22. Zhang M, Zhou M, Tang F, et al. Knowledge, attitude, and practice regarding COVID-19 among healthcare workers in Henan, China. J Hosp Infect. 2020;105:183-7.
23. Rajkumar RP. COVID-19 and mental health: a review of the existing literature. Asian J Psychiatr. 2020;52:102066.
24. Demirçelik Y, Kanık A. Covid-19 in children: epidemiology and clinical features. Anatol J Gen Med Res 2020;30:58-62.
25. Corbett GA, Milne SJ, Hehir MP, Lindow SW, O'connell MP. Health anxiety and behavioural changes of pregnant women during the COVID-19 pandemic. Eur J Obstet Gynecol Reprod Biol. 2020;249:96-7.
26. Republic of Turkey Ministry of Health. 14 rules against coronavirus risk [Internet]. 2020. Available from: http://www.istanbul.gov.tr/koronavirus-riskine-karsi-14-kural
27. World Health Organization. Coronavirus disease (COVID-19) advice for the public [Internet]. 2020. Available from: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/advice-for-public
28. World Health Organization. Coronavirus disease (COVID-19) outbreak: rights, roles and responsibilities of health workers, including key considerations for occupational safety and health [Internet]. 2020. Available from: https://www.who.int/docs/default-source/coronaviruse/who-rights-roles-respon-hw-covid-19.pdf
29. Asmundson GJG, Taylor S. How health anxiety influences responses to viral outbreaks like COVID-19: what all decision-makers, health authorities, and health care professionals need to know. J Anxiety Disord. 2020;71:102211.
30. Berthelot N, Lemieux R, Garon-Bissonnette J, Drouin-Maziade C, Martel É, Maziade M. Uptrend in distress and psychiatric symptomatology in pregnant women during the coronavirus disease 2019 pandemic. Acta Obstet Gynecol Scand. 2020;99:848-55.
31. Saccone G, Florio A, Aiello F, et al. Psychological impact of coronavirus disease 2019 in pregnant women. Am J Obstet Gynecol. 2020;223:293-5.
32. Wang C, Pan R, Wan X, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health. 2020;17:1729.

DOI: 10.4274/anatoljmed.2026.73645
Anatol J Gen Med Res 2026;36(2):161-70

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Performance of Large Language Models in Differentiating High-grade Gliomas and Solitary Brain Metastases

Yüksek Dereceli Gliomlar ile Soliter Beyin Metastazlarının Ayırt Edilmesinde Büyük Dil Modellerinin Performansı

Raşit Eren Büyüktoka[1], Ali Salbas[2], Miran Nihal Otlu[2], Ali Murat Koc[2], Asli Dilara Buyuktoka[2], Aslı Kahraman[3]

[1]University of Health Sciences Türkiye, İzmir City Hospital, Department of Radiology, İzmir, Türkiye
[2]İzmir Katip Çelebi University, İzmir Atatürk Training Research Hospital, Department of Radiology, İzmir, Türkiye
[3]İzmir Katip Çelebi University, İzmir Atatürk Training Research Hospital, Department of Pathology, İzmir, Türkiye

Cite as: Büyüktoka RE, Salbas A, Otlu MN, Koc AM, Buyuktoka AD, Kahraman A. Performance of large language models in differentiating high-grade gliomas and solitary brain metastases. Anatol J Gen Med Res. 2026;36(2):161-70

Abstract

Objective: To evaluate the diagnostic capability of three general-purpose multimodal large language models (LLMs) in differentiating high-grade gliomas from solitary brain metastases (SBMs) on conventional magnetic resonance imaging (MRI) and to compare their performance with radiologists of different expertise.

Methods: In this single-center retrospective study, 109 patients with pathologically confirmed supratentorial high-grade gliomas (n=67) or SBMs (n=42) were included. Axial T2-weighted, pre-contrast T1-weighted, and post-contrast T1-weighted images (single slice per sequence) were interpreted by three LLMs (ChatGPT-4.5, Gemini 2.5 Pro, Claude 4 Opus) and by four radiologists under different reading conditions. Accuracy, precision, recall, and weighted F1-scores were calculated; differences were tested with Cochran's Q and McNemar tests.

Results: Among all readers, diagnostic accuracy ranged from 31.19% to 77.06% across different reading conditions. With all sequences combined, the neuroradiologist achieved the highest accuracy (77.06%); the best-performing LLM was Claude 4 Opus. Combined-sequence evaluation improved accuracy compared with the best single-sequence performance for the neuroradiologist (from 72.48% to 77.06%) and for Claude 4 Opus (from 31.19% on pre-contrast T1WI to 66.06%).

Conclusion: General-purpose LLMs show potential in differentiating high-grade gliomas from SBMs on conventional MRI, but they remain inferior to experienced radiologists. Performance improves when multiple sequences are provided, but the error rate remains high. Future advancements in model architectures and volumetric data integration may enhance their clinical utility.

Keywords: Large language models, glioma, brain neoplasms, magnetic resonance imaging, neuroimaging

Address for Correspondence/Yazışma Adresi: Raşit Eren Büyüktoka, MD, University of Health Sciences Türkiye, İzmir City Hospital, Department of Radiology, İzmir, Türkiye
E-mail: rasiterenbuyuktoka@hotmail.com
ORCID ID: orcid.org/0000-0002-6989-6613

Received/Geliş tarihi: 27.01.2026
Accepted/Kabul tarihi: 23.03.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

Amaç: Bu çalışmanın amacı, konvansiyonel manyetik rezonans görüntüleme (MRG) kullanılarak yüksek dereceli gliomlar ile soliter beyin metastazlarının ayırt edilmesinde üç genel amaçlı çok modlu büyük dil modelinin (LLM) tanısal performansını değerlendirmek ve bu performansı farklı deneyim düzeylerine sahip radyologlarla karşılaştırmaktır.

Yöntem: Bu tek merkezli retrospektif çalışmaya, patolojik olarak doğrulanmış supratentoryal yüksek dereceli gliom (n=67) veya soliter beyin metastazı (n=42) tanısı bulunan toplam 109 hasta dahil edildi. Aksiyel T2 ağırlıklı, kontrast öncesi T1 ağırlıklı ve kontrast sonrası T1 ağırlıklı görüntüler (her sekans için tek kesit) üç LLM (ChatGPT-4.5, Gemini 2.5 Pro, Claude 4 Opus) ve farklı okuma koşullarında dört radyolog tarafından değerlendirildi. Doğruluk, kesinlik, duyarlılık ve ağırlıklı F1 skorları hesaplandı; gruplar arası farklar Cochran Q ve McNemar testleri ile analiz edildi.

Bulgular: Tüm değerlendiriciler arasında tanısal doğruluk, farklı okuma koşullarına bağlı olarak %31,19 ile %77,06 arasında değişti. Tüm sekanslar birlikte değerlendirildiğinde, en yüksek doğruluk oranına nöroradyolog ulaştı (%77,06); en iyi performans gösteren LLM ise Claude 4 Opus oldu. Birden fazla sekansın birlikte değerlendirilmesi, en iyi tek sekans performansına kıyasla hem nöroradyolog için (%72,48'den %77,06'ya) hem de Claude 4 Opus için (kontrast öncesi T1 ağırlıklı görüntülerde %31,19'dan %66,06'ya) doğruluğu artırdı.

Sonuç: Genel amaçlı LLMs, konvansiyonel MRG üzerinde yüksek dereceli gliomlar ile soliter beyin metastazlarını ayırt etmede potansiyel göstermektedir; ancak deneyimli radyologların performansının gerisinde kalmaktadır. Birden fazla sekansın birlikte sunulması performansı artırsa da, hata oranları halen yüksektir. Model mimarilerindeki gelişmeler ve volumetrik verilerin entegrasyonu, gelecekte bu modellerin klinik kullanım değerini artırabilir.

Anahtar Kelimeler: Büyük dil modelleri, gliom, beyin neoplazileri, manyetik rezonans görüntüleme, nörogörüntüleme

Introduction

Gliomas and brain metastases are the most frequent neoplastic diseases in neuro-oncology and are associated with high morbidity and mortality rates. Gliomas account for approximately 75% of all primary brain tumors in adults. In contrast, brain metastases represent the most frequently encountered malignant brain neoplasms overall, occurring at a rate nearly ten times higher than that of primary malignant brain tumors[1,2].

Accurate differentiation between solitary brain metastases (SBMs) and high-grade gliomas is important because each entity necessitates distinct therapeutic approaches[3,4]. Because their clinical presentations often overlap, this distinction typically relies on imaging findings. However, the conventional magnetic resonance imaging (MRI) appearances of SBMs and high-grade gliomas can be similar, and accurate early classification is pivotal because of differences in treatment. From a morphologic standpoint, gliomas, especially high-grade gliomas, grow infiltratively, permeating adjacent white-matter tracts, whereas SBMs expand along a pushing border that displaces surrounding parenchyma[5-7]. Multiparametric MRI protocols, particularly those quantitative perfusion metrics such as relative cerebral blood volume (rCBV), improve the differentiation of high-grade gliomas from SBMs; however, access to these advanced techniques remains constrained in many clinical settings[8-10].

The use of artificial intelligence (AI) applications in radiology has increased rapidly over the past decade. However, AI applications, such as deep learning algorithms, require large, carefully curated datasets with diverse and unbiased case distributions; this dependency limits the development of generalizable, high-quality models[11,12]. In contrast, publicly available large language models (LLMs) and their vision-language derivatives have demonstrated capabilities in relatively simple radiological works[13-19].

Whether multimodal LLMs can harness their broad visual-textual priors to distinguish SBMs from high-grade gliomas on conventional MRI remains unexplored. Understanding their performance in complex tasks, such as differentiating SBMs from high-grade gliomas, is critical, as success could enable the integration of automated image interpretation with natural language explanations, reduce the time from diagnosis to treatment, and facilitate rapid, cross-institutional deployment. This study is the first to evaluate the performance of multimodal LLMs in differentiating high-grade gliomas from SBMs.

Accordingly, this study evaluates the diagnostic accuracy and reliability of three publicly available closed-source multimodal LLMs in differentiating SBMs from high-grade gliomas on conventional MRI and benchmarks their performance against radiologists with varying levels of expertise. The primary objective of this study is to evaluate whether state-of-the-art multimodal LLMs can reliably discriminate SBMs from high-grade gliomas using conventional MRI and whether their diagnostic performance surpasses that of radiologists across levels of expertise.

Materials and Methods

This study was conducted as a single-center retrospective analysis. All methods performed in this study were in accordance with national and local laws, institutional guidelines, and ethical standards. The study was approved by İzmir Katip Çelebi University Institutional Ethics Committee (approval no: 0395, date: 19.06.2025). Informed consent was waived by the committee due to the design of the study.

The study design is illustrated in Figure 1. We identified a cohort of patients from our institution's pathology database who had pathologically confirmed supratentorial high-grade gliomas or solitary supratentorial brain metastases. These patients were reviewed to identify those who had undergone cranial MRI within one week prior to their surgical intervention. The primary inclusion criterion for this study was the presence of a solitary, newly diagnosed supratentorial intracranial lesion.

All relevant MRI examinations were retrieved from the institution's picture archiving and communication system. For each included patient, three axial image sets were extracted: T2-weighted, pre-contrast T1-weighted, and post-contrast T1-weighted. Fluid-attenuated inversion recovery (FLAIR) and diffusion weighted imaging were not analyzed because of heterogeneous imaging planes and incomplete coverage.

From each sequence, the single axial slice that demonstrated the most central portion of the lesion and had the largest diameter was selected for analysis. To ensure standardization across the dataset, all selected images were resized to 1024x1024 pixels. Prior to analysis, all patient-identifying information was removed from the images to ensure complete anonymization. The finalized case list was then randomized to create a unique order for interpretation.

Three distinct LLMs (ChatGPT 4.5, Google Gemini 2.5 Pro, and Claude 4-Opus) were used for image analysis. A single prompt was developed and used for all evaluations performed by the LLMs to ensure consistency and comparability of the results. The LLMs analyzed the images under the same conditions as the human readers by evaluating each of the three sequences individually and all sequences combined.

Four radiologists with varying levels of experience [a junior resident (M.N.O.) with 3 years of radiology experience, a senior resident (A.D.B.) with 5 years of radiology experience, a general radiologist (A.S.), and a subspecialist neuroradiologist (A.M.K.)] independently interpreted the images. Image interpretation was performed in four separate reading sessions: (1) T2-weighted images alone, (2) pre-contrast T1-weighted images alone, (3) post-contrast T1-weighted images alone, and (4) all three sequences combined.

To mitigate recall bias, a minimum washout period of two weeks was enforced between consecutive reading sessions for each radiologist. Furthermore, the order of the patient cases for each sequence was re-randomized before each reading session.

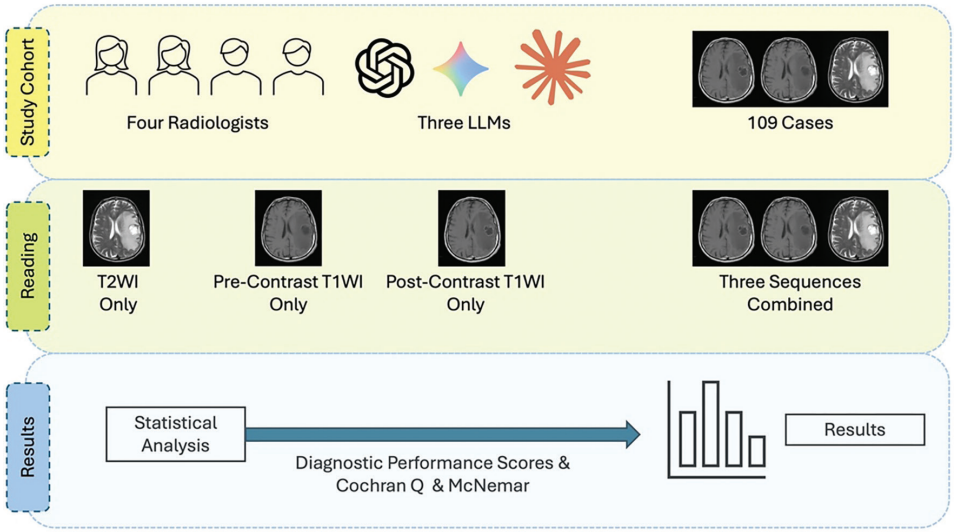


Figure 1. Overview of the study design

LLMs: Large language models

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics for Windows, version 22.0 (IBM Corp., Armonk, NY, USA).

Descriptive variables included age, maximum lesion diameter, gender, and World Health Organization grade. The diagnostic performance of each interpreter was evaluated by calculating accuracy, precision, recall, and weighted F1-score. To ensure robust estimation, 95% confidence intervals for each of these metrics were generated using a 1.000-repetition bootstrapping procedure. The diagnostic accuracy of seven interpreters (three LLMs and four radiologists) was compared. Thus, failure to detect a lesion was scored as a diagnostic error, regardless of the ground-truth class. To assess the overall difference in diagnostic accuracy across all interpreters for each imaging condition, Cochran's Q test was employed. In cases where the Cochran's Q test yielded a statistically significant result, post-hoc pairwise comparisons were conducted using the McNemar test. A Bonferroni correction was applied to the significance level for these multiple comparisons to control for type I error.

A p-value of less than 0.05 was considered to indicate statistical significance for all analyses, except where analyses were adjusted for multiple comparisons.

Results

The final study cohort comprised 109 patients. The patient selection process is depicted in the study flowchart (Figure 1).

The median age of the cohort was 58 years (interquartile range, 49-65 years). When stratified by ground-truth diagnosis, 67 (61.47%) cases were categorized as high-grade glial tumors and 42 (38.53%) cases were categorized as SBMs. Baseline demographic and clinical characteristics of the final study cohort are summarized in Table 1.

Diagnostic performance scores for each of the seven interpreters (3 LLMs and 4 radiologists) were summarized in Table 2, and Figures 2 and 3.

During analysis of T2-weighted images, LLMs failed to identify some lesions. Specifically, ChatGPT-4.5 did not identify a lesion in 3 patients (2.75%), Gemini 2.5 Pro did not identify a lesion in 6 patients (5.50%), and Claude 4 Opus did not identify a lesion in 10 patients (9.17%). In contrast, all radiologists successfully identified lesions in every case.

Table 1. Baseline demographic and clinical characteristics of patients with high-grade glioma and solitary brain metastases

Diagnosis	n (%)	Gender	Age (mean ± SD)	Maximum diameter (mm) [median (Q1-Q3)]
High-grade glioma	67 (61.47%)	Male=43 Female=14	53.06±15.24	44.0 (33.5-57.5)
Metastasis	43 (38.53%)	Male=30 Female=13	61.02±10.90	32.5 (24.25-45.75)

SD: Standard deviation

Table 2. Diagnostic performance metrics for each reader across magnetic resonance imaging sequence types. For each reader-sequence pairing, accuracy, precision, recall, and F1 score are reported with their 95% confidence intervals (CIs)

Reader	Sequence type	Accuracy (95% CIs)	Precision (95% CIs)	Recall (95% CIs)	F1 score (95% CIs)
ChatGPT 4.5	Combined	0.58 (0.49-0.66)	0.68 (0.57-0.77)	0.58 (0.49-0.66)	0.57 (0.47-0.66)
	Post-contrast T1WI	0.58 (0.49-0.67)	0.67 (0.57-0.77)	0.58 (0.49-0.67)	0.59 (0.49-0.68)
	Pre-contrast T1WI	0.52 (0.43-0.61)	0.63 (0.53-0.74)	0.52 (0.43-0.61)	0.56 (0.46-0.66)
	T2WI	0.61 (0.51-0.70)	0.59 (0.49-0.71)	0.61 (0.51-0.70)	0.58 (0.47-0.68)

Table 2. Continued

Reader	Sequence type	Accuracy (95% CIs)	Precision (95% CIs)	Recall (95% CIs)	F1 score (95% CIs)
Claude 4 Opus	Combined	0.66 (0.57-0.75)	0.65 (0.56-0.75)	0.66 (0.57-0.75)	0.65 (0.55-0.75)
	Post-contrast T1WI	0.61 (0.51-0.69)	0.61 (0.51-0.71)	0.61 (0.51-0.69)	0.60 (0.50-0.69)
	Pre-contrast T1WI	0.31 (0.23-0.39)	0.58 (0.46-0.70)	0.31 (0.23-0.39)	0.41 (0.30-0.49)
	T2WI	0.52 (0.42-0.61)	0.54 (0.42-0.66)	0.52 (0.42-0.61)	0.51 (0.41-0.61)
Gemini 2.5 Pro	Combined	0.58 (0.49-0.67)	0.58 (0.48-0.68)	0.58 (0.49-0.67)	0.58 (0.48-0.67)
	Post-contrast T1WI	0.58 (0.49-0.67)	0.59 (0.49-0.69)	0.58 (0.49-0.67)	0.58 (0.49-0.67)
	Pre-contrast T1WI	0.48 (0.39-0.57)	0.50 (0.40-0.61)	0.48 (0.39-0.57)	0.49 (0.39-0.58)
	T2WI	0.54 (0.45-0.63)	0.53 (0.43-0.63)	0.54 (0.45-0.63)	0.53 (0.43-0.62)
Neuroradiologist	Combined	0.77 (0.70-0.84)	0.79 (0.71-0.86)	0.77 (0.70-0.84)	0.77 (0.70-0.84)
	Post-contrast T1WI	0.72 (0.64-0.81)	0.74 (0.66-0.83)	0.72 (0.64-0.81)	0.73 (0.65-0.81)
	Pre-contrast T1WI	0.61 (0.51-0.81)	0.66 (0.56-0.75)	0.61 (0.51-0.71)	0.62 (0.53-0.71)
	T2WI	0.72 (0.63-0.80)	0.75 (0.67-0.83)	0.72 (0.63-0.80)	0.72 (0.64-0.80)
General radiologist	Combined	0.74 (0.66-0.82)	0.77 (0.70-0.85)	0.74 (0.66-0.82)	0.75 (0.67-0.82)
	Post-contrast T1WI	0.67 (0.58-0.75)	0.71 (0.63-0.80)	0.67 (0.58-0.75)	0.67 (0.59-0.76)
	Pre-contrast T1WI	0.58 (0.49-0.67)	0.61 (0.52-0.71)	0.58 (0.49-0.67)	0.58 (0.49-0.67)
	T2WI	0.71 (0.62-0.79)	0.74 (0.67-0.82)	0.71 (0.62-0.79)	0.71 (0.63-0.79)
Senior resident	Combined	0.72 (0.64-0.81)	0.74 (0.66-0.82)	0.72 (0.64-0.81)	0.73 (0.64-0.81)
	Post-contrast T1WI	0.67 (0.58-0.76)	0.71 (0.61-0.79)	0.67 (0.58-0.76)	0.67 (0.58-0.76)
	Pre-contrast T1WI	0.56 (0.47-0.65)	0.58 (0.49-0.68)	0.56 (0.47-0.65)	0.57 (0.48-0.65)
	T2WI	0.68 (0.60-0.76)	0.71 (0.63-0.80)	0.68 (0.60-0.76)	0.68 (0.60-0.77)

Table 2. Continued

Reader	Sequence type	Accuracy (95% CIs)	Precision (95% CIs)	Recall (95% CIs)	F1 score (95% CIs)
Junior resident	Combined	0.67 (0.57-0.75)	0.66 (0.56-0.75)	0.67 (0.57-0.75)	0.66 (0.55-0.75)
	Post-contrast T1WI	0.66 (0.57-0.74)	0.65 (0.55-0.74)	0.66 (0.57-0.74)	0.65 (0.55-0.73)
	Pre-contrast T1WI	0.64 (0.56-0.73)	0.65 (0.56-0.74)	0.64 (0.56-0.73)	0.64 (0.56-0.73)
	T2WI	0.74 (0.66-0.83)	0.75 (0.67-0.83)	0.74 (0.66-0.83)	0.74 (0.66-0.83)

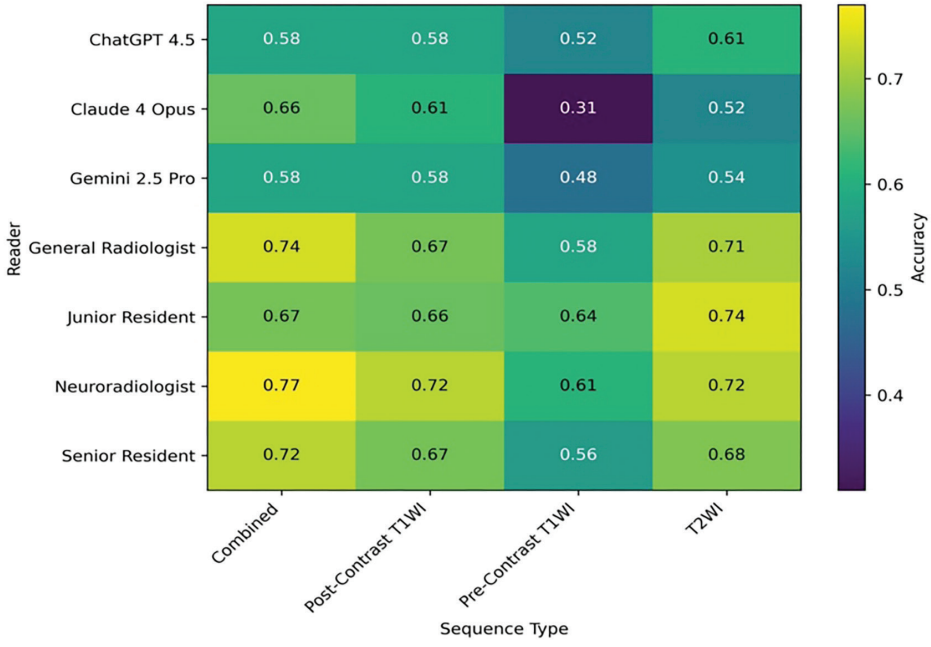


Figure 2. Accuracy heatmap by readers and sequence type. Rows represent the individual readers, including large language models and human radiologists. Columns represent reading types: combined (all sequences provided), post-contrast T1WI, pre-contrast T1WI, and T2WI. Color intensity reflects accuracy values, with lighter shades indicating lower accuracy and darker shades indicating higher accuracy. Numerical values in each cell denote the exact accuracy

Diagnostic accuracy rates for each of the seven readers on T2-weighted images ranged from 52.29% to 71.56%. A statistically significant difference in the correct classification rates among the interpreters was found (p=0.002). However, in subsequent pairwise comparisons, no statistically significant differences were identified between any pair of interpreters (a significance threshold of p<0.0024 was applied for multiple comparisons; all pairs of interpreters had p<0.0024).

LLMs exhibited a higher frequency of non-detected lesions when evaluating pre-contrast T1-weighted images. ChatGPT-4.5 failed to identify lesions in 21 patients (19.27%), Gemini 2.5 Pro failed to identify lesions in 9 patients (8.26%), and Claude 4 Opus failed to identify lesions in 50 patients (45.87%). The diagnostic accuracy of the readers ranged from 31.19% to 61.47%. A significant difference in performance among the interpreters was identified (p<0.001). Post-hoc pairwise comparisons revealed statistically significant differences between the performance of Claude 4 Opus and that of the neuroradiologist (p<0.001),

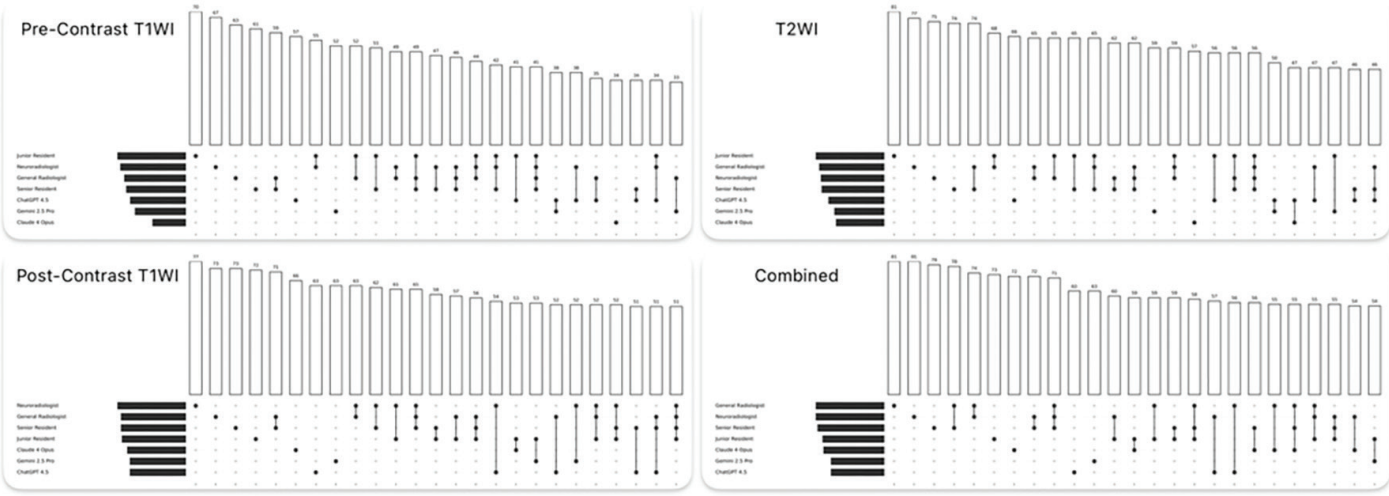


Figure 3. UpSet plot illustrating the overlap of correct classifications across readers and magnetic resonance imaging protocols. Each set corresponds to an individual reader (ChatGPT-4.5, Gemini 2.5 Pro, Claude 4 Opus, junior resident, senior resident, general radiologist, neuroradiologist), and protocol suffixes indicate the evaluated sequences (T2-weighted, pre-contrast T1WI, post-contrast T1WI, combined). Vertical bars represent the number of cases correctly classified in each specific intersection of readers, while filled dots below denote the readers included in that intersection. Horizontal bars display the total number of correct classifications per reader

the general radiologist (p<0.001), and the senior resident (p<0.001). Claude 4 Opus was inferior in diagnostic accuracy to all radiologists. A significant difference was also observed between ChatGPT-4.5 and Claude 4 Opus (p=0.001). No other pairwise comparisons yielded statistically significant results.

For post-contrast T1-weighted images, ChatGPT-4.5, Gemini 2.5 Pro, and Claude 4 Opus failed to identify a lesion in 6 (5.50%), 2 (1.83%), and 3 (2.75%) cases, respectively. Diagnostic performance, expressed as accuracy rates, ranged from 57.80% to 72.48%. A statistically significant difference was found among the interpreters (p=0.037). Nevertheless, subsequent post-hoc pairwise analysis revealed no significant differences among the interpreters (all p>0.0024, Bonferroni threshold for significance). In the analysis using the complete set of MRI sequences, ChatGPT-4.5 failed to identify a lesion in a single case (0.92%), whereas Gemini 2.5 Pro and Claude 4 Opus identified lesions in all cases. Diagnostic accuracy ranged from 57.80% to 77.06%. A statistically significant difference in performance was observed among the readers (p<0.001). Pairwise comparisons demonstrated that the neuroradiologist's performance was significantly superior to that of ChatGPT-4.5 (p<0.001) and Gemini 2.5 Pro (p=0.002). No other pairwise comparisons were statistically significant.

Discussion

In this study, we benchmarked the diagnostic performance of three multimodal LLMs against radiologists of varying expertise in the critical task of differentiating SBMs from high-grade gliomas. Our principal finding is that, while LLMs demonstrate some capability in this classification task, their performance is surpassed by expert human interpreters; a subspecialist neuroradiologist achieved the highest performance, with an accuracy score of 0.77. The best-performing LLM is Claude 4 Opus, which, when provided with all images combined, achieved an accuracy score of 0.66. Except for ChatGPT-4.5, all readers individually perform at their best across all sequences presented during the reading session. ChatGPT-4.5 performs best in a reading session when only T2WI is provided, achieving an accuracy score of 0.61. Another pivotal and concerning finding is the frequent failure of LLMs to identify lesions, particularly on pre-contrast T1WI. Claude 4 Opus missed lesions in 50 cases (45.87%).

The differentiation of high-grade gliomas from SBMs on conventional MRI presents a significant diagnostic challenge due to their overlapping morphological features. While advanced techniques such as perfusion imaging with rCBV can improve diagnostic accuracy, their application is not universal across all clinical settings[1-6]. With the advent of AI,

numerous studies have explored computational methods to distinguish brain metastases from glial tumors, particularly glioblastomas[20-24]. Many of these investigations have relied on radiomics, an approach that necessitates segmentation by expert radiologists for feature extraction. Other studies employing deep learning models, such as the work by Park et al.[24], have demonstrated promising results, achieving an area under the curve (AUC) of 0.83.

To our knowledge, there is no similar study assessing general-purpose LLMs' capabilities for such a complex binary classification task in neuro-oncology. Several attempts have been made to evaluate the performance of LLMs in challenging neuroradiology cases. For instance, Horiuchi et al.[25] compared LLMs with radiologists on complex cases and found that GPT-4 and GPT-4V achieved final diagnostic accuracies of only 22% and 16%, respectively. These results fell short of the performance of both radiology residents, who had accuracy rates between 28% and 31%, and board-certified radiologists, whose accuracy ranged from 38% to 47%[25]. Another investigation evaluating ChatGPT-4o, Grok, and Gemini for brain MRI interpretation found that for pathology prediction, accuracies of different pathologies such as gliomas, metastasis and cavernomas etc. ranged between 13.8% to 57.7%[26]. A separate analysis of ChatGPT-4o's diagnostic ability in diagnosing brain tumors highlighted a disparity between feature recognition and actual diagnosis. While it demonstrated high accuracy in identifying features such as perilesional edema (81%) and contrast enhancement (82.2%), its accuracy for determining the single most likely diagnosis was only 29.5%. This was significantly lower than the 65.9% to 70.5% accuracy achieved by radiologists in the same task[13]. These studies collectively suggest that while LLMs show promise in identifying specific imaging features, they are not yet able to match the diagnostic accuracy of human experts in complex neuroradiological tasks. Similarly, our results suggest that without domain-specific tuning or volumetric context, general-purpose multimodal LLMs do not yet surpass experienced radiologists on this task and may introduce a distinct failure mode.

Another important finding in our study is the failure of LLMs to detect lesions. Claude 4 Opus missed lesions in 50 cases (45.87%). Performance patterns reflected the information content of sequences with respect to this differential. Post-contrast T1WI usually has ring-enhancement and nodular margins typical of metastases and an enhancing tumor core in high-grade gliomas, while T2WI depicts infiltrative T2 hyperintensity and peritumoral edema patterns[6]. LLMs' higher non-detection rates occurred mostly on pre-contrast T1WI, which might be due to the nature of the tumor feature on pre-contrast T1-weighted images.

In the current study, it has been observed that the diagnostic performance of both human readers and LLMs, including ChatGPT-4.5 and a junior radiology resident, improves when a combined set of multiparametric MRI images is provided, as opposed to single sequences. This underscores the importance of a holistic assessment. Specifically, the neuroradiologist's accuracy increased from 72% with single sequences to 77% with combined images, and Claude 4 Opus's accuracy rose from 31% to 66%.

Similar to our findings, prior studies have demonstrated that combining multiple imaging sequences or views can substantially enhance classification performance in various radiological tasks. However, to our knowledge, no prior study has examined whether providing additional MRI sequences enhances the diagnostic performance of general-purpose LLMs. The only similar work to date has involved domain-specific, fine-tuned, or task-specific models. Bai et al.[27] demonstrated with their 3D multimodal large language model that incorporating full volumetric context across spatial perspectives markedly improved performance in visual question answering, retrieval, and segmentation tasks compared with 2D analysis. Furthermore, Kong et al.[28] reported that integrating T1-CE, T2, and T2-fluid-FLAIR sequences in a 3D ResNet-18 framework increased accuracy for glioblastoma vs. solitary brain metastasis discrimination to 87.2%, outperforming any single sequence. Likewise, in chest radiography, Hashir et al.[29] observed that adding lateral views to posteroanterior chest X-rays improved the AUC for various diagnostic labels, with gains comparable to doubling the training set size when using posteroanterior images alone. Collectively, these results suggest that the integration of multi-sequence and multi-view data provides richer contextual and spatial cues that benefit both deep learning architectures and emerging multimodal LLMs, although the magnitude of improvement may vary across models and levels of reader expertise.

Study Limitations

This study has several limitations that should be acknowledged. First, it was conducted at a single center and had a retrospective design, which may limit the generalizability of the findings to other institutions with different patient populations, MRI protocols, or scanner hardware. Second, only three conventional MRI sequences

(axial T2-weighted, axial pre-contrast T1-weighted, and axial post-contrast T1-weighted) and a single representative slice per sequence were provided for analysis. This approach omits volumetric information, advanced sequences, such as perfusion imaging, and multiplanar reconstructions that may carry diagnostic cues. Third, only closed-source, general-purpose, multimodal LLMs were evaluated; the performance of open-source, domain-specific, fine-tuned medical models or of prompt engineering techniques such as few-shot learning was not assessed, and results may differ in such settings. Finally, although lesion detection failures were documented, the closed-source nature of these models meant we could not perform a detailed qualitative error analysis of the underlying failure modes in LLM decision-making, limiting insights into how these models process and prioritize imaging features. Despite these limitations, this study is the first to compare LLMs and radiologists in differentiating high-grade gliomas from SBMs, providing a novel contribution to the literature.

Future research should enhance general-purpose multimodal LLMs by incorporating detection-first pipelines that localize lesions and extract features before classification. While it is not currently possible to provide full 3D volumetric inputs to the general-purpose LLMs used in this study, advances in model architectures may eventually allow multi-slice or volumetric analysis, preserving critical spatial information lost in single-slice evaluation. Future directions might include prompt optimization and fine-tuning of open-source models such as large language-and-vision assistant[30].

Conclusion

In this head-to-head comparison, general-purpose multimodal LLMs demonstrated a measurable, albeit inferior, ability to differentiate high-grade gliomas from SBMs on conventional MRI compared with radiologists, particularly those with subspecialty neuroradiology expertise. While performance improved when multiple MRI sequences were provided, LLMs nonetheless failed because lesions were frequently not detected. These findings underscore that in their current form, LLMs cannot replace expert radiological interpretation for this challenging neuro-oncologic task. However, their potential to augment clinical workflows remains, particularly if future developments enable volumetric image processing, domain-specific fine-tuning, and integration of lesion detection pipelines. Such advancements could narrow the performance gap and support broader, cost-effective deployment of LLM-assisted diagnostic tools in neuroimaging.

Ethics

Ethics Committee Approval: The study was approved by İzmir Katip Çelebi University Institutional Ethics Committee (approval no: 0395, date: 19.06.2025).

Informed Consent: Informed consent was waived by the committee due to the design of the study.

Footnotes

Authorship Contributions

Concept: R.E.B., Design: R.E.B., A.S., A.D.B., Data Collection or Processing: A.S., M.N.O., A.M.K., A.D.B., A.K., Analysis or Interpretation: R.E.B., A.S., Literature Search: R.E.B., Writing: R.E.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. Lamba N, Wen PY, Aizer AA. Epidemiology of brain metastases and leptomeningeal disease. Neuro Oncol. 2021;23:1447-56.
2. Lapointe S, Perry A, Butowski NA. Primary brain tumours in adults. Lancet. 2018;392:432-46.
3. Weller M, van den Bent M, Preusser M, et al. EANO guidelines on the diagnosis and treatment of diffuse gliomas of adulthood. Nat Rev Clin Oncol. 2021;18:170-86.
4. Vogelbaum MA, Brown PD, Messersmith H, et al. Treatment for brain metastases: ASCO-SNO-ASTRO Guideline. J Clin Oncol. 2022;40:492-516.
5. Suh CH, Kim HS, Jung SC, Kim SJ. Diffusion-weighted imaging and diffusion tensor imaging for differentiating high-grade glioma from solitary brain metastasis: a systematic review and meta-analysis. AJNR Am J Neuroradiol. 2018;39:1208-14.
6. Jung BY, Lee EJ, Bae JM, Choi YJ, Lee EK, Kim DB. Differentiation between glioblastoma and solitary metastasis: morphologic assessment by conventional brain MR imaging and diffusion-weighted imaging. Investig Magn Reson Imaging. 2021;25:23-34.
7. Blanchet L, Krooshof PW, Postma GJ, et al. Discrimination between metastasis and glioblastoma multiforme based on morphometric analysis of MR images. AJNR Am J Neuroradiol. 2011;32:67-73.
8. Suh CH, Kim HS, Jung SC, Choi CG, Kim SJ. Perfusion MRI as a diagnostic biomarker for differentiating glioma from brain metastasis: a systematic review and meta-analysis. Eur Radiol. 2018;28:3819-31.
9. Mouthuy N, Cosnard G, Abarca-Quinones J, Michoux N. Multiparametric magnetic resonance imaging to differentiate high-grade gliomas and brain metastases. J Neuroradiol. 2012;39:301-7.

10. Abreu VS, Tarrio J, Silva J, et al. Multiparametric analysis from dynamic susceptibility contrast-enhanced perfusion MRI to evaluate malignant brain tumors. J Neuroimaging. 2024;34:257-66.
11. Tripathi S, Gabriel K, Dheer S, et al. Understanding biases and disparities in radiology AI datasets: a review. J Am Coll Radiol. 2023;20:836-41.
12. Büyüktoka RE, Adibelli ZH, Sürücü M, et al. Magnetic resonance imaging-based artificial intelligence model predicts neoadjuvant therapy response in triple-negative breast cancer. Diagn Interv Radiol. 2025.
13. Ozenbas C, Engin D, Altinok T, Akcay E, Aktas U, Tabanli A. ChatGPT-4o's performance in brain tumor diagnosis and MRI Findings: a comparative analysis with radiologists. Acad Radiol. 2025;32:3608-17.
14. Bhayana R. Chatbots and large language models in radiology: a practical primer for clinical and research applications. Radiology. 2024;310:e232756.
15. Akinci D'Antonoli T, Stanzione A, Bluethgen C, et al. Large language models in radiology: fundamentals, applications, ethical considerations, risks, and future directions. Diagn Interv Radiol. 2024;30:80-90.
16. Elek A, Ekizalioğlu DD, Güler E. Evaluating microsoft bing with ChatGPT-4 for the assessment of abdominal computed tomography and magnetic resonance images. Diagn Interv Radiol. 2025;31:196-205.
17. Çamur E, Cesur T, Güneş YC, et al. Evaluating large language models for imaging modality selection: potential to reduce unnecessary contrast agent use and radiation exposure. Clin Imaging. 2025;125:110573.
18. Singh R, Hamouda M, Chamberlin JH, et al. ChatGPT vs. Gemini: comparative accuracy and efficiency in lung-RADS score assignment from radiology reports. Clin Imaging. 2025;121:110455.
19. Xia X, Wu W, Tan Q, Gou Q. Interpretable machine learning models for differentiating glioblastoma from solitary brain metastasis using radiomics. Acad Radiol. 2025;32:5388-400.
20. Büyüktoka RE, Salbas A. Multimodal large language models for pediatric bone-age assessment: a comparative accuracy analysis. Acad Radiol. 2025;32:6905-12.
21. Qian Z, Li Y, Wang Y, et al. Differentiation of glioblastoma from solitary brain metastases using radiomic machine-learning classifiers. Cancer Lett. 2019;451:128-35.
22. Artzi M, Bressler I, Ben Bashat D. Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. J Magn Reson Imaging. 2019;50:519-28.
23. Li Y, Liu Y, Liang Y, et al. Radiomics can differentiate high-grade glioma from brain metastasis: a systematic review and meta-analysis. Eur Radiol. 2022;32:8039-51.
24. Park YW, Eom S, Kim S, et al. Differentiation of glioblastoma from solitary brain metastasis using deep ensembles: empirical estimation of uncertainty for clinical reliability. Comput Methods Programs Biomed. 2024;254:108288.
25. Horiuchi D, Tatekawa H, Oura T, et al. Comparing the diagnostic performance of GPT-4-based ChatGPT, GPT-4V-based ChatGPT, and radiologists in challenging neuroradiology cases. Clin Neuroradiol. 2024;34:779-87.
26. Sozer A, Sahin MC, Sozer B, et al. Do LLMs have 'the eye' for MRI? Evaluating GPT-4o, Grok, and Gemini on brain MRI performance: First Evaluation of Grok in medical imaging and a comparative analysis. Diagnostics (Basel). 2025;15:1320.
27. Bai F, Du Y, Huang T, Meng MQH, Zhao B. M3D: advancing 3D medical image analysis with multi-modal large language models. arXiv. 2024.
28. Kong C, Yan D, Liu K, Yin Y, Ma C. Multiple deep learning models based on MRI images in discriminating glioblastoma from solitary brain metastases: a multicentre study. BMC Med Imaging. 2025;25:171.
29. Hashir M, Bertrand H, Cohen JP. Quantifying the value of lateral views in deep learning for chest X-rays. arXiv . 2020.
30. Liu H, Li C, Wu Q, Lee YJ. Visual instruction tuning. arXiv. 2023.

DOI: 10.4274/anatoljmed.2026.77598
Anatol J Gen Med Res 2026;36(2):171-6

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Sarcopenia and its Association with the Timing of Metastasis in Breast Cancer: A Retrospective Cohort Study

Metastatik Meme Kanserinde Sarkopeninin Metastaz Zamanı ile İlişkisi: Retrospektif Kohort Çalışması

Erkan Güler, Mahsum Barçin

Mersin University Faculty of Medicine, Department of Surgery, Division of Surgical Oncology, Mersin, Türkiye

Cite as: Güler E, Barçin M. Sarcopenia and its association with the timing of metastasis in breast cancer: a retrospective cohort study. Anatol J Gen Med Res. 2026;36(2):171-6

Abstract

Objective: Sarcopenia is described as a progressive reduction that is increasingly recognized as a factor that can influence clinical outcomes in cancer patients. However, in metastatic breast cancer, its clinical significance remains unclear, particularly regarding when metastatic disease develops. This study therefore examines whether sarcopenia is related to the timing of metastasis and influences overall survival.

Methods: This study screened 530 female patients who were diagnosed with breast cancer between 2010 and 2020. Among them, 60 patients who developed metastatic disease and had available computed tomography imaging suitable for body composition analysis were included. Patients were categorized as having metastasis at initial diagnosis or during follow-up. Sarcopenia was quantified by measuring the psoas muscle area at the level of the L3 vertebra and normalizing these measurements to patient height. Survival outcomes were analyzed with Kaplan-Meier curves.

Results: Sarcopenia was present in 12 patients (20.0%). It was observed more often among patients who had metastatic disease at the time of diagnosis than among those who developed metastasis later during follow-up (42.9% vs. 13.0%, p=0.024). Over a median follow-up of 29.4 months, mortality was higher in the sarcopenic group (50.0% vs. 12.5%, p=0.009). Despite this, no significant difference was found in overall survival when analyzed using time-to-event methods.

Conclusion: Sarcopenia may be more common among patients presenting with metastatic disease and may be associated with higher mortality in metastatic breast cancer. However, its effect on overall survival is still unclear. Further evidence from larger, prospectively designed studies is required to better define its clinical relevance.

Keywords: Metastatic breast cancer, psoas muscle index, sarcopenia

Öz

Amaç: Sarkopeni, iskelet kas kütlesinde ilerleyici azalma ile karakterize olup kanser hastalarında sonuçları etkileyebileceği öne sürülmektedir. Metastatik meme kanserinde rolü, özellikle metastatik hastalığın zamanlaması ile ilişkisi açısından net değildir. Bu çalışmanın amacı, sarkopeni ile metastaz zamanlaması arasındaki ilişkiyi ve genel sağkalım üzerindeki olası etkisini değerlendirmektir.

Yöntem: Bu retrospektif çalışmada 2010-2020 yılları arasında meme kanseri tanısı almış 530 kadın hasta tarandı. Bunlardan metastatik hastalık gelişen ve vücut kompozisyon analizi için uygun bilgisayarlı tomografi görüntüleri bulunan 60 hasta çalışmaya dahil edildi. Hastalar metastazın tanı anında

Address for Correspondence/Yazışma Adresi: Asst. Prof., Erkan Güler, Mersin University Faculty of Medicine, Department of Surgery, Division of Surgical Oncology, Mersin, Türkiye
E-mail: drerkangler@gmail.com
ORCID ID: orcid.org/0000-0003-0881-1003

Received/Geliş tarihi: 13.04.2026
Accepted/Kabul tarihi: 05.05.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

mevcut olması veya takip sırasında gelişmesine göre gruplandırıldı. Sarkopeni, üçüncü lomber vertebra düzeyinde ölçülen psoas kas alanlarının boya göre normalize edilmesi ile değerlendirildi. Sağkalım analizi Kaplan-Meier yöntemi ile yapıldı.

Bulgular: Sarkopeni 12 hastada (%20,0) saptandı. Tanı anında metastatik hastalığı olan hastalarda, takip sırasında metastaz gelişen hastalara göre daha sık bulundu (%42,9'a karşı %13,0; p=0,024). Medyan 29,4 aylık takip süresince sarkopenisi olan hastalarda mortalite daha yüksek bulundu (%50,0'a karşı %12,5; p=0,009). Ancak gruplar arasında genel sağkalım açısından istatistiksel olarak fark yoktu.

Sonuç: Sarkopeni, metastatik meme kanserinde hastalığın başlangıçtaki sunumu ve artmış mortalite ile ilişkili görünmektedir. Bununla birlikte genel sağkalım üzerindeki etkisi net olarak gösterilememiştir. Bu konuda geniş ölçekli prospektif çalışmalar gereklidir.

Anahtar Kelimeler: Metastatik meme kanseri, psoas kas indeksi, sarkopeni

Introduction

Loss of muscle mass or function-often termed sarcopenia-may influence patient outcomes. Prior research has linked sarcopenia to higher rates of morbidity and mortality across a range of solid tumors, suggesting that it may reflect diminished physiological reserve and a tendency toward less favorable oncological outcomes[1].

There has been growing interest in how body composition affects clinical outcomes in breast cancer. In particular, alterations in skeletal muscle mass have been linked to treatment tolerance and survival, especially in advanced disease settings. Metastatic breast cancer represents a biologically heterogeneous condition with variable clinical behavior and treatment responses, as emphasized in international consensus guidelines[2]. Therefore, identifying host-related factors that may influence disease course and outcomes remains of clinical importance.

Recent evidence suggests that sarcopenia may also have prognostic implications in metastatic breast cancer. Studies have reported that impaired muscle parameters, such as low muscle attenuation, are associated with poorer survival in patients receiving systemic therapy[3]. Recent cohort studies suggest that the presence of sarcopenia at baseline may adversely affect outcomes in certain subgroups of metastatic breast cancer, including patients receiving CDK4/6 inhibitor therapy[4]. Similarly, findings from a recent systematic review and meta-analysis indicate that sarcopenia is linked to higher mortality and disease progression in breast cancer, reinforcing its potential clinical importance[5]. Despite this growing body of evidence, the role of sarcopenia in metastatic breast cancer is not yet fully understood. In particular, the relationship between sarcopenia and the timing of metastatic disease-whether metastasis is present at initial diagnosis or develops during follow-up-has not been clearly elucidated. Moreover, data regarding the impact of sarcopenia on survival outcomes in this specific context remain limited.

In this context, we examined whether sarcopenia-defined using the psoas muscle index (PMI)-is related to the timing of metastatic disease and whether it influences overall survival for patients with metastatic breast cancer.

Materials and Methods

Study Design and Patient Selection

This study retrospectively reviewed the records of patients treated at Mersin University Faculty of Medicine. Between January 2010 and December 2020, 530 individuals with a diagnosis of breast cancer were evaluated. Of these, 60 patients who developed metastatic disease and had imaging data appropriate for body composition assessment were ultimately included.

Patients were categorized into two groups according to the timing of metastatic disease: those with metastatic disease at initial diagnosis and those who developed metastasis during follow-up (secondary metastatic disease).

Ethical Approval

Ethical approval was obtained from the Ethics Committee of Mersin University (approval no: 2026/110, date: 04.03.2026). All procedures were carried out in line with the principles outlined in the Declaration of Helsinki.

Data Collection

Clinical and pathological information was retrospectively obtained from the institutional electronic records. The dataset included age at diagnosis, histopathological subtype, and hormone receptor profile. The timing of metastatic

disease was determined based on imaging and clinical records. The date of diagnosis was defined as the date of the first positron emission tomography-computed tomography (CT) examination.

Assessment of Sarcopenia

Sarcopenia was evaluated using the PMI, derived from axial contrast-enhanced CT images obtained at the level of the third lumbar vertebra (L3), a widely used reference point for skeletal muscle assessment. The right and left psoas muscles were manually outlined at the L3 level, and the measured areas were adjusted for patient height squared (mm^2/m^2). PMI was determined using the following formula:

PMI (mm^2/m^2)= (left psoas muscle area+right psoas muscle area)/$height^2$

All measurements were performed on pre-treatment CT images using a standardized anatomical reference. All PMI measurements were performed by a single experienced observer using a standardized measurement protocol based on predefined anatomical landmarks at the L3 level. To minimize measurement variability, all assessments were conducted using the same methodology under consistent conditions. Formal intra- and interobserver reproducibility analyses were not performed because measurements were conducted by a single observer; this is acknowledged as a limitation. Sex-specific cut-off values derived from previously published studies were applied to define sarcopenia. Because the study cohort consisted exclusively of female patients, sarcopenia was defined using the previously published female-specific PMI threshold of <360 mm^2/m^2, which has been used in oncologic populations and validated in prior studies[6].

Outcome Measures

We examined whether sarcopenia is associated with the timing of metastatic disease onset. Overall survival was assessed and defined as the time from diagnosis to death from any cause or to the last available follow-up.

Statistical Analysis

Discrete variables are summarized as counts with corresponding percentages and are compared between groups using Fisher's exact test. Numerical data are described either by the mean and standard deviation or by the median and interquartile range, depending on their distribution.

Time-to-event outcomes were analyzed using Kaplan-Meier estimation, and group differences were assessed using the log-rank test. Statistical significance was set at a two-sided p-value below 0.05. All analyses were conducted using conventional statistical software.

Results

Patient Selection and Baseline Characteristics

Between January 2010 and December 2020, 530 patients with a diagnosis of breast cancer were evaluated. Among these, 60 patients who developed metastatic disease and had available imaging data suitable for body composition analysis were included in the final analysis. Of the 60 patients, 14 (23.3%) presented with metastatic disease at initial diagnosis, while 46 (76.7%) developed metastatic disease during follow-up. The mean age of the cohort was 53.6±12.8 years, with no significant difference between the *de novo* and secondary metastatic groups (p=0.86). Similarly, the estrogen receptor and human epidermal growth factor receptor 2 statuses were comparable between the two groups (p=0.49 and p=0.29, respectively). Baseline clinicopathological characteristics, stratified by timing of metastatic disease, are summarized in Table 1.

Association Between Sarcopenia and Timing of Metastatic Disease

Sarcopenia, as defined by PMI, was identified in 12 patients (20.0%) within the metastatic cohort. Among patients presenting with *de novo* metastatic disease, 6 of 14 (42.9%) were classified as sarcopenic. In contrast, among patients who developed metastasis during follow-up, 6 of 46 (13.0%) were sarcopenic. This difference was statistically significant (p=0.024), indicating a potential association between sarcopenia and *de novo* metastatic presentation. The distribution of sarcopenia according to metastatic timing is presented in Table 2.

Survival Outcomes

A mean follow-up of 31.5 months (median: 29.4 months), 12 patients (20.0%) died. Mortality was more frequent among patients with sarcopenia than among those without (50.0% vs. 12.5%, p=0.009). Even with this difference, there was no significant difference for overall survival between two groups (log-rank p>0.05), and median overall survival was not reached in either group. Survival data are summarized in Table 3.

Table 1. Baseline clinicopathologic characteristics according to timing of metastatic disease

Variable	Total (n=60)	*De novo* metastatic (n=14)	Metastasis during follow-up (n=46)	p-value
Age, years (mean ± SD)	53.6±12.8	53.1±12.1	53.8±13.0	0.86
Sarcopenia, n (%)	12 (20.0)	6 (42.9)	6 (13.0)	0.024
Non-sarcopenia, n (%)	48 (80.0)	8 (57.1)	40 (87.0)	
ER positive, n (%)	40 (66.7)	8 (57.1)	32 (69.6)	0.49
HER2 positive, n (%)	5 (8.3)	0 (0.0)	5 (10.9)	0.29

SD: Standard deviation, HER2: Human epidermal growth factor receptor 2, ER: Estrogen receptor

Table 2. Comparison of clinicopathologic characteristics according to sarcopenia status

Variable	Total (n=60)	Sarcopenic (n=12)	Non-sarcopenic (n=48)	p-value
Age, years (mean ± SD)	53.6±12.8	55.2±11.5	53.2±13.1	0.62
De novo metastasis, n (%)	14 (23.3)	6 (50.0)	8 (16.7)	0.024
Metastasis during follow-up, n (%)	46 (76.7)	6 (50.0)	40 (83.3)	
ER positive, n (%)	40 (66.7)	7 (58.3)	33 (68.8)	0.52
HER2 positive, n (%)	5 (8.3)	1 (8.3)	4 (8.3)	1.00
Death during follow-up, n (%)	12 (20.0)	6 (50.0)	6 (12.5)	0.009

Values are reported as mean ± standard deviation (SD) or as number (%). Fisher's exact test was used to calculate p-values HER2: Human epidermal growth factor receptor 2, ER: Estrogen receptor

Table 3. Overall survival outcomes according to sarcopenia status

Variable	Sarcopenic (n=12)	Non-sarcopenic (n=48)	p-value
Number of deaths, n (%)	6 (50.0)	6 (12.5)	0.009
Median follow-up, months	29.4	29.4	-
Mean follow-up, months	31.5	31.5	-
Median overall survival	Not reached	Not reached	-
Log-rank test	-	-	>0.05

Values are given as numbers (%) unless stated otherwise. Median overall survival was not reached in either group

Discussion

In this study including 60 patients metastatic breast cancer, sarcopenia-defined by the PMI-was identified in 12 patients (20.0%). Sarcopenia was significantly more frequent in patients with *de novo* metastatic disease than in those who developed metastasis during follow-up (42.9% vs. 13.0%, p=0.024). In addition, the proportion of deaths during follow-up was higher among sarcopenic patients than in non-sarcopenic patients (50.0% vs. 12.5%, p=0.009). Kaplan-Meier analysis did not show difference between 2 groups for overall survival. These findings suggest that sarcopenia may be associated with higher mortality, while its effect on overall survival (time-to-event) in this cohort remains uncertain, possibly because the limited number of events reduced the ability to detect differences in survival despite observed differences in crude mortality rates.

The observed association between sarcopenia and *de novo* metastatic presentation may reflect the broader clinical and biological significance of muscle depletion in breast cancer. Recent evidence suggests that sarcopenia is not merely a consequence of advanced disease but may represent an underlying host-related condition associated with more aggressive tumor behavior. In a analysis focusing on metastatic breast cancer, sarcopenia was associated with adverse clinical outcomes and poorer prognosis[7]. Similarly, Deluche et al.[8] highlighted that sarcopenia may be an under-recognized factor in the management of metastatic breast cancer, emphasizing its potential role in influencing disease course and patient outcomes. Taken together, this suggests that sarcopenia may be linked to an unfavorable disease phenotype and could contribute to the early manifestation of metastatic disease observed in our cohort.

The association between sarcopenia and survival has been widely reported, although the extent of its effect appears to differ across study populations. Evidence from a recent systematic review and meta-analysis indicates that sarcopenia is linked to increased mortality in patients with breast cancer, supporting its role as a negative prognostic factor[9]. Likewise, another meta-analysis focusing on muscle composition found that patients with sarcopenia experienced higher mortality, earlier disease progression, and greater treatment-related toxicity[10].

In the present cohort, mortality was numerically higher among patients with sarcopenia; however, this difference was not significance, may be attributable to the limited number of observed events. These observations suggest that although sarcopenia is generally associated with poorer survival, its effect may be less apparent in smaller datasets where statistical power is constrained.

The absence of a significant difference in overall survival between patients with and without sarcopenia in this cohort may reflect the complex, multifactorial determinants of prognosis in metastatic breast cancer. Clinical outcomes are influenced not only by muscle mass but also by a range of host-related factors, including inflammatory status and nutritional status. In line with this, previous studies have suggested that combining body composition parameters with systemic inflammatory markers provides a more comprehensive understanding of survival in metastatic breast cancer[11]. Reductions in muscle-related indices have been associated with not only survival outcomes but also with tumor progression dynamics. For example, lower pectoralis muscle index has been reported as an independent predictor of both distant metastasis-free survival and overall survival in breast cancer populations[12]. These observations indicate that muscle depletion should be interpreted within a broader biological context rather than as an isolated parameter and that PMI alone may not fully capture the interaction between host physiology and tumor behavior. The lack of a detectable effect on overall survival in the present study may therefore be related to methodological limitations, particularly the relatively small sample size and the limited number of events, which may have reduced the statistical power of the analyses. This should be considered when interpreting the lack of a significant difference in survival.

Study Limitations

The findings of this study have some limitations. The retrospective, single-center design might have led to selection bias. A small sample size and a limited number of survival events may reduce statistical power, particularly for overall survival analyses. Third, sarcopenia assessment was based solely on PMI and did not include functional measures of muscle performance. Measurements were performed by a single observer without formal reproducibility analysis.

Conclusion

In conclusion, our findings suggest that sarcopenia, as assessed by PMI, may be associated with *de novo* metastatic presentation and increased mortality in metastatic breast cancer. Despite these findings, a clear effect on overall survival could not be established. Further prospective studies with larger cohorts are required to better define its prognostic and clinical relevance.

Ethics

Ethics Committee Approval: Ethical approval was obtained from the Ethics Committee of Mersin University (approval no: 2026/110, date: 04.03.2026).

Informed Consent: Retrospective study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: E.G., Concept: E.G., M.B., Design: E.G., M.B., Data Collection or Processing: E.G., M.B., Analysis or Interpretation: E.G., M.B., Literature Search: E.G., M.B., Writing: E.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.

References

1. Prado CM, Lieffers JR, McCargar LJ, et al. Prevalence and clinical implications of sarcopenic obesity in patients with solid tumours of the respiratory and gastrointestinal tracts: a population-based study. Lancet Oncol. 2008;9:629-35.
2. Cardoso F, Senkus E, Costa A, et al. 4th ESO-ESMO International Consensus Guidelines for advanced breast cancer (ABC 4)†. Ann Oncol. 2018;29:1634-57.
3. Rier HN, Jager A, Sleijfer S, van Rosmalen J, Kock MCJM, Levin MD. Low muscle attenuation is a prognostic factor for survival in metastatic breast cancer patients treated with first line palliative chemotherapy. Breast. 2017;31:9-15.
4. Karahan L, Akyildiz A, Sahin TK, et al. Prognostic value of baseline sarcopenia and adipose tissue indices in HR+/HER2- metastatic breast

cancer treated with CDK4/6 inhibitors: a retrospective cohort study. J Clin Med. 2026;15:1623.

5. Roberto M, Barchiesi G, Resuli B, et al. Sarcopenia in breast cancer patients: a systematic review and meta-analysis. Cancers (Basel). 2024;16:596.
6. Güler E, Gümüş T, Sayur V, et al. Preoperative psoas muscle index predicts long-term survival but not postoperative morbidity after curative-intent gastrectomy for gastric cancer. Int J Gen Med. 2026;19:581421.
7. Jang MK, Park S, Park C, Raszewski R, Park S, Kim S. Sarcopenia in patients with metastatic breast cancer: a systematic review and meta-analysis. Breast. 2025;82:104508.
8. Deluche E, Lachatre D, Di Palma M, et al. Is sarcopenia a missed factor in the management of patients with metastatic breast cancer? Breast. 2022;61:84-9.
9. Dai Y, Lan J, Li S, Xu G. Exploring the impact of sarcopenia on mortality in breast cancer patients: a comprehensive systematic review and meta-analysis. Breast Care (Basel). 2024;19:316-28.
10. Aleixo GFP, Williams GR, Nyrop KA, Muss HB, Shachar SS. Muscle composition and outcomes in patients with breast cancer: meta-analysis and systematic review. Breast Cancer Res Treat. 2019;177:569-79.
11. Yamanouchi K, Murakami S, Sato A, Ogawa S, Shinagawa H, Kamohara Y. Integrated evaluation of inflammatory, nutritional, and sarcopenia markers to predict survival in metastatic breast cancer patients. In Vivo. 2023;37:811-7.
12. Huang WJ, Zhang ML, Wang W, et al. Preoperative pectoralis muscle index predicts distant metastasis-free survival in breast cancer patients. Front Oncol. 2022;12:854137.

DOI: 10.4274/anatoljmed.2026.70188
Anatol J Gen Med Res 2026;36(2):177-83

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Prognostic Factors Affecting Survival in Thyroid Carcinoma: A Retrospective Single-center Study

Tiroid Karsinomunda Sağkalımı Etkileyen Prognostik Faktörler: Retrospektif Tek Merkezli Bir Çalışma

Semra Salimoğlu, Ayberk Dursun, Ekrem Kocatürk, Ragıp Kayar

University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of General Surgery, İzmir, Türkiye

Cite as: Salimoğlu S, Dursun A, Kocatürk E, Kayar R. Prognostic factors affecting survival in thyroid carcinoma: a retrospective single-center study. Anatol J Gen Med Res. 2026;36(2):177-83

Abstract

Objective: To investigate the relationship between clinical, pathological, and treatment-related characteristics and survival outcomes in patients with thyroid carcinoma.

Methods: This retrospective study included 45 patients with histopathologically confirmed thyroid carcinoma who were treated and followed at the Surgical Department of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital over a 20-year period. Demographic features, histological tumor subtypes, metastatic status, and surgical treatment details were reviewed. Potential prognostic factors associated with survival were evaluated.

Results: The mean age of the patients was 48±15 years. Of the study population, 71.1% were female and 28.9% were male. Papillary thyroid carcinoma was the predominant histological subtype, representing 60% of all cases. Lymph node metastasis was present in 35.6% of patients, whereas distant metastasis was present in 17.8% of patients. Patients with differentiated thyroid carcinomas showed better survival outcomes than those with other histological subtypes. Older age, metastatic disease, and aggressive tumor histology were associated with reduced survival. Among these factors, distant metastasis showed the most pronounced association with mortality.

Conclusion: Differentiated thyroid carcinomas are generally associated with favorable survival; however, the presence of metastases and aggressive histological subtypes adversely affect prognosis. Age, histological subtype, and metastatic status should be considered important factors in prognostic assessment and risk stratification of patients with thyroid carcinoma.

Keywords: Thyroid cancer, differentiated thyroid carcinoma, metastasis, prognosis, survival

Address for Correspondence/Yazışma Adresi: Semra Salimoğlu, MD, University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, Department of General Surgery, İzmir, Türkiye
E-mail: semrasalimoglu@yahoo.co.uk
ORCID ID: orcid.org/0000-0002-8793-1603

Received/Geliş tarihi: 19.04.2026
Accepted/Kabul tarihi: 04.06.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

Amaç: Bu çalışmada, tiroid karsinomu tanısıyla takip ve tedavi edilen hastalarda klinik ve histopatolojik özelliklerin sağkalım üzerine olan etkisinin değerlendirilmesi amaçlandı.

Yöntem: Retrospektif olarak planlanan bu çalışmaya, Sağlık Bilimleri Üniversitesi, İzmir Tepecik Eğitim ve Araştırma Hastanesi Cerrahi Kliniği'nde 20 yıllık süreç içerisinde takip edilen ve histopatolojik inceleme sonucunda tiroid karsinomu tanısı kesinleşen 45 hasta dahil edildi. Hastalara ait demografik veriler, tümörün histolojik alt tipi, lenf nodu ve uzak metastaz durumu ile uygulanan cerrahi tedavi yöntemleri incelendi. Sağkalım üzerinde etkili olabilecek klinik ve patolojik prognostik faktörler değerlendirildi.

Bulgular: Çalışmaya dahil edilen hastaların yaş ortalaması 48±15 yıl idi. Olguların %71,1'i kadın, %28,9'u erkekti. Histolojik alt tipler arasında en sık papiller tiroid karsinomu saptandı ve bu grup tüm olguların %60'ını oluşturdu. Lenf nodu metastazı hastaların %35,6'sında, uzak metastaz ise %17,8'inde mevcuttu. Diferansiye tiroid karsinomu olan hastalarda sağkalım oranlarının daha yüksek olduğu görüldü. Buna karşılık ileri yaş, metastatik hastalık varlığı ve agresif histolojik alt tipler daha kötü sağkalım sonuçları ile ilişkiliydi. Uzak metastaz varlığı, mortalite ile en güçlü ilişki gösteren faktör olarak öne çıktı.

Sonuç: Diferansiye tiroid karsinomlarında sağkalım genel olarak iyi olmakla birlikte, metastatik hastalık ve agresif histolojik alt tipler prognozu olumsuz yönde etkilemektedir. Bu nedenle tiroid kanseri tanılı hastalarda yaş, histolojik alt tip ve metastaz durumunun prognozun belirlenmesinde dikkate alınması gereken temel faktörler olduğu düşünülmektedir.

Anahtar Kelimeler: Tiroid kanseri, diferansiye tiroid karsinomu, metastaz, prognoz, sağkalım

Introduction

Thyroid carcinoma is the leading malignant disease of the endocrine system and represents a relatively small proportion of all cancers. Over recent decades, the number of diagnosed thyroid cancer cases has increased in many regions. This increase is generally attributed to the broader use of neck ultrasonography and cross-sectional imaging, which has led to more frequent detection of small and clinically silent thyroid nodules. In contrast to the rise in incidence, mortality has remained relatively stable, particularly among patients with differentiated thyroid carcinoma, in whom long-term outcomes are usually favorable[1,2].

Histopathologically, thyroid malignancies are primarily classified as papillary, follicular, medullary, and anaplastic carcinomas. Papillary and follicular carcinomas are commonly evaluated together as differentiated thyroid carcinomas and account for the great majority of thyroid malignancies[3]. Papillary thyroid carcinoma is the predominant subtype and often has an indolent clinical course. By contrast, anaplastic carcinoma is rare but is associated with highly aggressive biological behavior and poor survival[4].

Despite the generally favorable course of differentiated thyroid carcinoma, the prognosis is not uniform across all patients. Some patients develop recurrence, metastatic spread, or disease-related mortality during follow-up. Therefore, defining prognostic variables is important for risk assessment, treatment planning, and long-term surveillance. Age, sex, tumor size, histological subtype, capsular invasion, extrathyroidal extension, nodal status, and distant metastasis have all been reported to influence survival outcomes in thyroid carcinoma[5].

Age is one of the most frequently used parameters in prognostic evaluation. Older patients have been shown to experience higher disease-specific mortality in several clinical series. Tumor-related characteristics, particularly larger tumor diameter and extrathyroidal extension, may also reflect more advanced disease and have been associated with poorer outcomes[6].

Regional lymph node metastasis is especially common in papillary thyroid carcinoma. Its effect on survival may vary depending on patient age, tumor burden, and other pathological features; however, nodal disease remains an important component of staging and risk evaluation. Distant metastasis more directly and adversely affects survival. In differentiated thyroid carcinoma, the lungs and bones are among the most frequent metastatic sites, and the development of distant metastatic disease is associated with a substantial reduction in long-term survival[7,8].

Several scoring systems have been developed to estimate prognosis in thyroid cancer. Models such as AGES, AMES, and MACIS combine clinical and pathological parameters including age, metastasis, tumor extent, tumor size, grade, completeness of resection, and invasion. These systems may help clinicians stratify patients into risk groups and choose appropriate follow-up and treatment strategies[9].

Although most thyroid cancers follow a favorable clinical course, outcomes may differ considerably between patients. For this reason, recognition of high-risk features and individualized risk stratification remain essential in the management of thyroid carcinoma. Evaluation of demographic, clinical, and pathological characteristics may provide useful information for estimating prognosis and tailoring follow-up after treatment[10].

The aim of this study was to evaluate the relationship between clinical and pathological characteristics and survival outcomes in patients with thyroid carcinoma treated at a single tertiary center over a 20-year period.

Materials and Methods

Study Design and Patient Population

This retrospective study was designed to assess factors associated with survival in patients with thyroid carcinoma. Patients who underwent surgical treatment and who were followed in the Department of Surgery at University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital were evaluated. The study cohort consisted of 45 patients with thyroid malignancy who were treated during a 20-year period. Demographic data, clinical presentation, laboratory findings, operative details, pathological results, and follow-up information were retrieved from hospital archives and patient files.

The diagnosis of thyroid carcinoma was confirmed in all cases by histopathological examination of surgical specimens. The study was approved by University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. Data recorded at diagnosis, tumor-related characteristics, and clinical outcomes during follow-up were reviewed for each patient.

Inclusion and Exclusion Criteria

Patients were eligible for inclusion if they had histopathologically confirmed thyroid carcinoma, had undergone surgery at the Department of Surgery at University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, and had clinical, pathological, and follow-up data sufficient for survival assessment. Patients were excluded if the diagnosis was not confirmed histopathologically, essential clinical or pathological information was missing, follow-up data could not be obtained, or surgical treatment was not performed.

Clinical and Demographic Variables

The variables analyzed included age, sex, clinical presentation, tumor size at diagnosis, type of surgical treatment, lymph node metastasis, distant metastasis, extrathyroidal extension, and histopathological tumor subtype.

Laboratory data recorded during diagnosis and follow-up were reviewed retrospectively. These parameters included serum levels of thyroid-stimulating hormone, triiodothyronine, thyroxine, thyroglobulin, and calcitonin, with calcitonin particularly considered in patients with medullary thyroid carcinoma. Laboratory findings were evaluated together with the clinical data available at diagnosis.

Pathological Evaluation

Surgical specimens were assessed in the pathology laboratory. Tumors were classified according to histological features, and pathological variables, including tumor diameter, capsular invasion, extrathyroidal extension, and lymph node metastasis, were recorded when available.

Follow-up and Survival Assessment

Follow-up information was obtained from outpatient records and patient files. Survival duration was calculated from the date of diagnosis to the date of death or the last documented follow-up visit. Overall survival (OS) was defined as the primary endpoint.

Statistical Analysis

Statistical analysis was performed using SPSS for Windows (SPSS Inc., Chicago, IL, USA, version 7.5). Descriptive data were presented as mean ± standard deviation for continuous variables and as numbers and percentages for categorical variables. Categorical variables were compared using the chi-square test, whereas continuous variables were analyzed using the Student's t-test. Survival was evaluated using the Kaplan-Meier method, and differences between survival curves were assessed using the log-rank test. A p-value <0.05 was accepted as statistically significant.

Results

The study included 45 patients with histopathologically confirmed thyroid carcinoma who underwent treatment in the Department of Surgery at University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital, during a 20-year period. Demographic and clinical features of the cohort are presented in Table 1.

Patient age ranged between 18 and 80 years, and the mean age was 48±15 years. Twenty-seven patients (60%) were younger than 45 years, while 18 (40%) were 45 years of age or older.

There were 32 female patients (71.1%) and 13 male patients (28.9%), corresponding to a female-to-male ratio of approximately 2.5:1 (Table 1).

Papillary thyroid carcinoma was the most frequent histological diagnosis, identified in 27 patients (60%). Follicular carcinoma was observed in 9 patients (20%), and Hürthle cell carcinoma was observed in 4 patients (8.9%). Less common subtypes were medullary carcinoma (n=2, 4.4%), anaplastic carcinoma (n=1, 2.2%), and other malignancies (n=2, 4.4%).

Overall, differentiated thyroid carcinomas, including papillary and follicular subtypes, accounted for approximately 80% of the study population (Table 2).

Lymph node metastasis was present in 16 patients (35.6%), and distant metastasis was identified in 8 patients (17.8%). The lungs and bones were the most common sites of distant metastasis.

Among patients with papillary carcinoma, lymph node metastasis was detected in 10 (37%), whereas distant metastasis was present in 2 (7.4%). Among patients with follicular carcinoma, the rate of distant metastasis was 33.3% (Table 2).

All patients in the cohort underwent total thyroidectomy. Depending on disease extent, nodal involvement, and clinical risk profile, selected patients underwent additional radical procedures, such as modified neck dissection. Additional radical surgery was performed in a total of 5 patients: 4 in the high-risk group and 1 in the low-risk group. Patients classified as high-risk more commonly underwent more extensive surgical procedures (Table 3).

In the analysis of survival-related factors, lower survival rates were observed in patients of advanced age, male patients, and patients with metastatic disease. Survival was 96.3% in patients younger than 45 years and 77.8% in those aged 45 years or older. Mortality was also higher in male patients than in female patients (Table 4).

Metastatic status was closely related to survival outcomes. Survival was 96.6% in patients without lymph node metastasis, and 75.0% in those with nodal involvement.

Table 1. Demographic and clinical characteristics of the study population

Variable	n (%)
Age (mean ± SD)	48±15
Age range	18-80
Age groups	
<45 years	27 (60)
≥45 years	18 (40)
Sex	
Female	32 (71.1)
Male	13 (28.9)
Histological subtype	
Papillary carcinoma	27 (60)
Follicular carcinoma	9 (20)
Hürthle cell carcinoma	4 (8.9)
Medullary carcinoma	2 (4.4)
Anaplastic carcinoma	1 (2.2)
Other malignancies	2 (4.4)
Lymph node metastasis	16 (35.6)
Distant metastasis	8 (17.8)

SD: Standard deviation

Table 2. Clinical course and oncological outcomes according to histological subtypes

Histological subtype	Patients, n (%)	LN metastasis, n (%)	Distant metastasis, n (%)	Recurrence, n (%)	Death, n (%)
Papillary carcinoma	27 (60)	10 (37)	2 (7.4)	3 (11.1)	1 (3.7)
Follicular carcinoma	9 (20)	3 (33.3)	3 (33.3)	0 (0)	1 (11.1)
Hürthle cell carcinoma	4 (8.9)	0 (0)	0 (0)	0 (0)	0 (0)
Medullary carcinoma	2 (4.4)	1 (50)	1 (50)	0 (0)	1 (50)
Anaplastic carcinoma	1 (2.2)	1 (100)	1 (100)	1 (100)	1 (100)
Other malignancies	2 (4.4)	1 (50)	1 (50)	1 (50)	1 (50)
Total	45 (100)	16 (35.6)	8 (17.8)	5 (11.1)	5 (11.1)

LN: Lymph node

The difference was more pronounced for distant metastasis: survival was 97.3% in patients without distant metastasis and 50.0% in patients with distant metastasis (Table 4).

Survival was highest among patients with differentiated thyroid carcinoma. The survival rates were 96.3% for papillary thyroid carcinoma and 88.9% for follicular carcinoma. In contrast, mortality was higher in medullary carcinoma and particularly evident in anaplastic thyroid carcinoma.

The duration of follow-up ranged from 1 month to 20 years, with a mean follow-up period of approximately 8 years. During follow-up, recurrence and disease-related death each occurred in 5 patients (11.1%).

Major prognostic factors and corresponding survival outcomes are summarized in Table 5. Patients aged ≥45 years had lower survival than younger patients. Similarly, nodal or distant metastatic disease was associated with reduced survival. Papillary and follicular carcinomas showed better survival outcomes than medullary and anaplastic carcinomas.

Discussion

In this retrospective single-center study, clinical and pathological factors associated with survival were evaluated in 45 patients treated for thyroid carcinoma over a 20-year period. The results showed that papillary thyroid

Table 3. Distribution of surgical procedures according to risk groups

Surgical treatment	Low-risk group, n (%)	High-risk group, n (%)	Total, n (%)
Total thyroidectomy	23 (100.0)	22 (100.0)	45 (100.0)
Additional radical surgical procedure	1 (4.3)	4 (18.2)	5 (11.1)
No additional radical surgical procedure	22 (95.7)	18 (81.8)	40 (88.9)

Table 4. Prognostic factors associated with survival

Prognostic factor	Patients, n (%)	Death, n (%)	Survival (%)
Age <45 years	27 (60)	1 (3.7)	96.3
Age ≥45 years	18 (40)	4 (22.2)	77.8
Female	32 (71.1)	2 (6.3)	93.7
Male	13 (28.9)	3 (23.1)	76.9
No lymph node metastasis	29 (64.4)	1 (3.4)	96.6
Lymph node metastasis present	16 (35.6)	4 (25)	75
No distant metastasis	37 (82.2)	1 (2.7)	97.3
Distant metastasis present	8 (17.8)	4 (50)	50

Table 5. Overall survival according to major prognostic factors

Prognostic factor	Category	Patients, (n)	Deaths, (n)	Survival (%)
Age	<45 years	27	1	96.3
Age	≥45 years	18	4	77.8
Sex	Female	32	2	93.7
Sex	Male	13	3	76.9
Lymph node metastasis	Absent	29	1	96.6
Lymph node metastasis	Present	16	4	75
Distant metastasis	Absent	37	1	97.3
Distant metastasis	Present	8	4	50
Histological subtype	Papillary carcinoma	27	1	96.3
Histological subtype	Follicular carcinoma	9	1	88.9
Histological subtype	Hürthle cell carcinoma	4	0	100
Histological subtype	Medullary carcinoma	2	1	50
Histological subtype	Anaplastic carcinoma	1	1	0

carcinoma was the most common histological subtype, that differentiated thyroid carcinomas had better survival outcomes, and that survival was lower among patients of advanced age, those with metastatic disease, and those with aggressive histological subtypes.

Papillary thyroid carcinoma accounted for approximately 60% of the cases in the present cohort, making it the predominant subtype. This distribution is in line with the general epidemiological pattern of thyroid malignancies, in which differentiated thyroid cancers, particularly papillary and follicular carcinomas, constitute the majority of cases[11]. Consistent with this biological profile, patients with papillary and follicular carcinomas in our cohort had higher survival rates than those with less common and more aggressive subtypes[12].

Age has long been incorporated into prognostic assessment in thyroid cancer. Classical staging and scoring systems have frequently used age thresholds, including 45 years, to estimate risk[13,14]. In our cohort, patients aged 45 years or older had lower survival rates than younger patients, supporting the clinical relevance of age for outcome evaluation.

However, age should not be interpreted as an isolated determinant of prognosis. Contemporary risk assessment emphasizes the combined evaluation of patient age, tumor biology, disease extent, and response to treatment. Therefore, age is best considered together with pathological and clinical risk factors rather than as a stand-alone parameter[15].

Histological subtype was another important factor related to the outcome. Survival was higher in differentiated tumors, including papillary and follicular carcinomas, whereas outcomes were worse in medullary and particularly in anaplastic thyroid carcinomas. Although anaplastic thyroid carcinoma is uncommon, it is known for its rapidly progressive course and poor prognosis[16]. These findings highlight the prognostic value of tumor histology in addition to its diagnostic role.

Regional lymph node metastases are frequently encountered in papillary thyroid carcinoma. Its prognostic significance has been debated, because some studies suggest a stronger association with locoregional recurrence than with OS[17]. In the present study, patients with lymph node metastasis had poorer survival than those without nodal disease. This observation may indicate that nodal involvement reflects a higher tumor burden or more aggressive disease characteristics in selected patients.

Distant metastasis was the factor most clearly associated with reduced survival in this cohort. Survival was markedly lower in patients with distant metastatic disease than in those without distant metastatic disease. This finding is compatible with previous reports showing that metastatic spread to distant organs substantially worsens prognosis, even in differentiated thyroid carcinoma[18]. The lungs and bones were the most frequent metastatic sites in our patients, which is consistent with the typical metastatic pattern of differentiated thyroid cancers[19].

Extrathyroidal extension and local invasion are also important components of risk evaluation in thyroid carcinoma. Previous studies have linked extrathyroidal extension with increased recurrence risk and mortality[20]. In the present study, these variables were assessed as part of the pathological evaluation and contributed to the overall characterization of disease extent and risk profile.

Risk stratification in thyroid cancer has evolved over time. In addition to traditional staging systems, dynamic risk stratification models that incorporate response to treatment are now widely used to refine prognosis during follow-up[21]. The long time span covered by the present study may limit direct comparisons with modern risk-adapted management strategies. Nevertheless, our findings support the ongoing relevance of classical prognostic variables such as age, histological subtype, and metastatic status.

Study Limitations

This study has some limitations. The small sample size reduced statistical power, particularly for analyses of rare histological subtypes. In addition, the retrospective design and the 20-year study period may have resulted in variations in diagnostic methods, treatment approaches, and follow-up practices. The long follow-up is a strength of the study, providing information about long-term outcomes in thyroid carcinoma.

The results indicate that prognosis in patients with thyroid carcinoma is influenced by both patient- and tumor-related factors. Differentiated thyroid carcinomas generally have favorable survival, whereas metastatic disease and aggressive histology are associated with poorer outcomes.

These findings support the importance of risk-based follow-up and individualized treatment planning.

Conclusion

This retrospective study evaluated the effect of clinical and pathological characteristics on survival in patients with thyroid carcinoma. Differentiated thyroid carcinomas were the most common histological group and were associated

with favorable survival outcomes. In contrast, advanced age, metastatic disease, and aggressive histological subtypes were associated with decreased survival.

Age, tumor histology, and metastatic status should be considered key parameters when estimating the prognosis for patients with thyroid carcinoma. Early recognition of these factors may help clinicians identify high-risk patients and plan appropriate follow-up strategies.

Differentiated thyroid carcinomas usually have favorable outcomes; however, metastasis and aggressive histological subtypes adversely affect prognosis. Risk-adapted management and long-term surveillance remain essential in the care of patients with thyroid carcinoma.

Ethics

Ethics Committee Approval: The study was approved by University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital.

Informed Consent: Retrospective study.

Footnotes

This study is based on the specialty thesis conducted by Dr. Semra Salimoğlu at University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital in 1999. At the time the study was conducted, an ethics committee and an ethics committee approval number system in their current form were not in place at our institution. Specialty thesis studies were conducted with the approval of the Education Planning Committee (EPC).

The thesis approval document issued by the EPC at that time has been submitted to the journal. Therefore, a separate ethics committee approval number is not available for this study.

Authorship Contributions

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

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

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

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74:12-49.
2. Cabanillas ME, McFadden DG, Durante C. Thyroid cancer. Lancet. 2016;388:2783-95.
3. Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines Task Force on thyroid nodules and differentiated thyroid cancer. Thyroid. 2016;26:1-133.
4. Fagin JA, Wells SA Jr. Biologic and clinical perspectives on thyroid cancer. N Engl J Med. 2016;375:1054-67.
5. Ito Y, Miyauchi A. Prognostic factors and therapeutic strategies for differentiated carcinomas of the thyroid. Endocr J. 2009;56:177-92.
6. Nixon IJ, Wang LY, Migliacci JC, et al. An international multi-institutional validation of age 55 years as a cutoff for risk stratification in the AJCC/UICC staging system for well-differentiated thyroid cancer. Thyroid. 2016;26:373-80.
7. Randolph GW, Duh QY, Heller KS, et al. The prognostic significance of nodal metastases from papillary thyroid carcinoma can be stratified based on the size and number of metastatic lymph nodes, as well as the presence of extranodal extension. Thyroid. 2012;22:1144-52.
8. Durante C, Haddy N, Baudin E, et al. Long-term outcome of 444 patients with distant metastases from papillary and follicular thyroid carcinoma: benefits and limits of radioiodine therapy. J Clin Endocrinol Metab. 2006;91:2892-9.
9. Hay ID, Bergstralh EJ, Goellner JR, Ebersold JR, Grant CS. Predicting outcome in papillary thyroid carcinoma: development of a reliable prognostic scoring system in a cohort of 1779 patients surgically treated at one institution during 1940 through 1989. Surgery. 1993;114:1050-7; discussion 1057-8.
10. Mazzaferri EL, Jhiang SM. Long-term impact of initial surgical and medical therapy on papillary and follicular thyroid cancer. Am J Med. 1994;97:418-28.
11. La Vecchia C, Malvezzi M, Bosetti C, et al. Thyroid cancer mortality and incidence: a global overview. Int J Cancer. 2015;136:2187-95.
12. Lim H, Devesa SS, Sosa JA, Check D, Kitahara CM. Trends in thyroid cancer incidence and mortality in the United States, 1974-2013. JAMA. 2017;317:1338-48.
13. Shah JP, Loree TR, Dharker D, Strong EW, Begg C, Vlamis V. Prognostic factors in differentiated carcinoma of the thyroid gland. Am J Surg. 1992;164:658-61.
14. Bellantone R, Lombardi CP, Boscherini M, et al. Prognostic factors in differentiated thyroid carcinoma: a multivariate analysis of 234 consecutive patients. J Surg Oncol. 1998;68:237-41.
15. Tuttle RM, Tala H, Shah J, et al. Estimating risk of recurrence in differentiated thyroid cancer after total thyroidectomy and radioactive iodine remnant ablation: using response to therapy variables to modify the initial risk estimates predicted by the new American Thyroid Association staging system. Thyroid. 2010;20:1341-9.
16. Smallridge RC, Ain KB, Asa SL, et al. American Thyroid Association guidelines for management of patients with anaplastic thyroid cancer. Thyroid. 2012;22:1104-39.
17. Dhutt GS, Ellis L, Wiseman SM. The prognostic significance of lymph node metastases characteristics in papillary thyroid carcinoma: a scoping review. Expert Rev Anticancer Ther. 2026;26:479-92.
18. Shaha AR, Shah JP, Loree TR. Differentiated thyroid cancer presenting initially with distant metastasis. Am J Surg. 1997;174:474-6.
19. Schlumberger M, Challeton C, De Vathaire F, et al. Radioactive iodine treatment and external radiotherapy for lung and bone metastases from thyroid carcinoma. J Nucl Med. 1996;37:598-605.
20. Radowsky JS, Howard RS, Burch HB, Stojadinovic A. Impact of degree of extrathyroidal extension of disease on papillary thyroid cancer outcome. Thyroid. 2014;24:241-4.
21. Lee SG, Lee WK, Lee HS, et al. Practical performance of the 2015 American Thyroid Association Guidelines for predicting tumor recurrence in patients with papillary thyroid cancer in South Korea. Thyroid. 2017;27:174-81.

DOI: 10.4274/anatoljmed.2026.49091
Anatol J Gen Med Res 2026;36(2):184-9

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

The Impact of Radical Nephrectomy on Local Recurrence and Prognosis in Primary Retroperitoneal Liposarcoma: A Comprehensive Analysis of Oncological and Physiological Outcomes

Primer Retroperitoneal Liposarkomda Radikal Nefrektominin Lokal Nüks ve Prognoz Üzerindeki Etkisi: Onkolojik ve Fizyolojik Sonuçların Kapsamlı Analizi

Berk Göktepe, Kamil Erözkan, Ebubekir Korucuk, Safa Vatansever, Tufan Gümüş, Alper Uğuz, Cemil Çalışkan, Osman Bozbıyık

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

Cite as: Göktepe B, Erözkan K, Korucuk E, et al. The impact of radical nephrectomy on local recurrence and prognosis in primary retroperitoneal liposarcoma: a comprehensive analysis of oncological and physiological outcomes. Anatol J Gen Med Res. 2026;36(2):184-9

Abstract

Objective: Primary retroperitoneal liposarcoma (RPLS) is a rare and biologically heterogeneous malignancy for which complete macroscopic resection remains the cornerstone of curative treatment. However, the role of en bloc radical nephrectomy in achieving optimal oncological outcomes remains controversial. This study aimed to evaluate the impact of radical nephrectomy on local recurrence, survival outcomes, and postoperative renal function in patients undergoing curative-intent resection for primary RPLS.

Methods: A retrospective cohort study was conducted, including patients who underwent complete surgical resection for primary RPLS at a tertiary referral center between January 2011 and January 2024. Patients were categorized according to surgical strategy into those undergoing en bloc nephrectomy and those undergoing kidney-sparing resection. The primary endpoint was local recurrence, while secondary endpoints included disease-free survival (DFS), overall survival (OS), perioperative morbidity, and postoperative renal outcomes, assessed using kidney disease: improving global outcomes 2024 criteria.

Results: A total of 21 patients (57.1% female) with a median age of 51 years were included (radical nephrectomy: 6; kidney-sparing surgery: 15). The median tumor size was 27 cm, and the majority of tumors were classified as undifferentiated (57.1%). The nephrectomy group demonstrated a lower rate of local recurrence and a longer median DFS than the kidney-sparing group. Five-year OS was also higher in patients undergoing nephrectomy. Importantly, in the nephrectomy cohort, no patient developed postoperative acute kidney injury, and no major complications were observed.

Conclusion: En bloc radical nephrectomy appears to improve local disease control and survival outcomes in selected patients with primary RPLS without compromising renal function. These findings support an aggressive surgical approach when required to achieve adequate oncological margins.

Keywords: Retroperitoneal liposarcoma, radical nephrectomy, kidney-sparing surgery, local recurrence, overall survival

Address for Correspondence/Yazışma Adresi: Kamil Erözkan, MD, Ege University Faculty of Medicine, Department of General Surgery, İzmir, Türkiye
E-mail: kamil.erozkan@ege.edu.tr
ORCID ID: orcid.org/0000-0003-2193-9984

Received/Geliş tarihi: 22.04.2026
Accepted/Kabul tarihi: 11.06.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

Amaç: Primer retroperitoneal liposarkom (RPLS), nadir görülen ve biyolojik açıdan heterojen bir malignitedir. Küratif tedavinin temelini makroskopik tam rezeksiyon oluşturmakla birlikte, en blok radikal nefrektominin optimal onkolojik sonuçlara katkısı halen tartışmalıdır. Bu çalışmada, radikal nefrektominin lokal nüks, sağkalım sonuçları ve postoperatif renal fonksiyon üzerindeki etkisinin değerlendirilmesi amaçlanmıştır.

Yöntem: Ocak 2011 ile Ocak 2024 tarihleri arasında üçüncü basamak bir merkezde primer RPLS nedeniyle küratif amaçlı cerrahi rezeksiyon uygulanan hastalar retrospektif olarak incelendi. Hastalar cerrahi yaklaşıma göre en blok nefrektomi uygulananlar ve böbrek koruyucu cerrahi yapılanlar olarak iki gruba ayrıldı. Birincil sonlanım noktası lokal nüks olup, ikincil sonlanım noktaları hastalıksız sağkalım (DFS), genel sağkalım (OS), perioperatif morbidite ve böbrek yetmezliği: global sonuçları iyileştirme 2024 kriterlerine göre değerlendirilen postoperatif renal fonksiyon sonuçları olarak belirlendi.

Bulgular: Ortanca yaşı 51 olan toplam 21 hasta (%57,1 kadın) çalışmaya dahil edildi (radikal nefrektomi: 6, böbrek koruyucu cerrahi: 15 hasta). Nefrektomi grubunda lokal nüks oranının daha düşük olduğu ve medyan DFS süresinin daha uzun olduğu gözlendi. Beş yıllık genel sağkalım oranı da nefrektomi uygulanan hastalarda daha yüksekti. Ayrıca, hiçbir hastada postoperatif akut böbrek hasarı gelişmedi ve nefrektomi grubunda majör komplikasyon izlenmedi.

Sonuç: Primer RPLS cerrahisinde en blok radikal nefrektomi, uygun hasta seçimi yapıldığında renal fonksiyonu olumsuz etkilemeden lokal hastalık kontrolünü ve sağkalım sonuçlarını iyileştirebilir. Bu bulgular, yeterli onkolojik sınırların sağlanması gereken olgularda daha agresif cerrahi yaklaşımın tercih edilebileceğini desteklemektedir.

Anahtar Kelimeler: Retroperitoneal liposarkom, radikal nefrektomi, böbrek koruyucu cerrahi, lokal nüks, genel sağkalım

Introduction

Soft tissue sarcomas are rare mesenchymal malignancies accounting for less than 1% of all adult cancers, with approximately 15% to 20% arising in the retroperitoneum[1]. Among retroperitoneal sarcomas, liposarcomas represent the most common histological subtype[1,2]. Due to the expansile capacity of the retroperitoneal space, these tumors frequently remain asymptomatic until they reach considerable size, often exceeding 20 cm in diameter[2].

The biological behavior of retroperitoneal liposarcoma is closely associated with the histological subtype[3]. Well-differentiated liposarcoma typically follows an indolent course but demonstrates a high propensity for local recurrence. In contrast, dedifferentiated (DDLPS) liposarcoma is characterized by more aggressive biological features, including higher proliferative activity and increased metastatic potential[4-6]. Despite these differences, complete surgical resection remains the only potentially curative treatment modality[7,8].

Achieving negative surgical margins in the retroperitoneum is technically challenging due to the absence of well-defined anatomical compartments[9]. As a result, extended resections involving adjacent organs are frequently required. Radical compartmental resection has been proposed to improve local control by removing contiguous structures that may harbor microscopic disease[10,11].

The necessity of nephrectomy in this context remains a matter of ongoing debate. While en bloc removal of the kidney may facilitate wider margins, concerns persist regarding postoperative renal impairment and its long-term consequences[12]. Recent consensus recommendations advocate for a selective, histology-driven approach, favoring organ preservation whenever oncologically feasible[13].

At the same time, the preservation of renal function has gained increasing importance in the era of multimodal therapy. Systemic treatments used in high-risk sarcomas may include nephrotoxic agents, making adequate renal reserve essential for eligibility and tolerability[14]. Therefore, surgical decision-making must carefully balance oncological radicality against functional preservation.

The present study aims to evaluate the impact of radical nephrectomy on local recurrence, survival outcomes, and renal function in patients undergoing resection for primary retroperitoneal liposarcoma (RPLS).

Materials and Methods

This retrospective observational study was conducted at a tertiary care center. Medical records of patients who underwent curative-intent surgery for RPLS between January 2011 and January 2024 were reviewed. Institutional ethical approval was obtained from Ege University prior to data collection (approval no: 24-12.1T/62, date: 26.12.2024).

Patients included in the study were adults with histopathologically confirmed RPLS who underwent complete macroscopic resection without tumor fragmentation. Patients with recurrent disease, non-liposarcoma histology,

distant metastases at diagnosis, or intraoperative tumor rupture were excluded.

All patients underwent preoperative staging using contrast-enhanced computed tomography or magnetic resonance imaging. Surgical procedures were performed through a midline laparotomy, aiming to achieve complete macroscopic resection with either negative or microscopically positive margins.

Patients were divided into two groups based on the intraoperative management of the ipsilateral kidney. In the nephrectomy group, en bloc resection of the tumor included removal of the ipsilateral kidney due to direct invasion or inability to safely separate the tumor from the kidney. In the kidney-sparing group, the tumor was resected while preserving the kidney.

Clinical and pathological data were collected, including demographic characteristics, comorbidities, operative details, and histological findings. Comorbidity burden was assessed using the Charlson comorbidity index, and physiological status was evaluated using the American Society of Anesthesiologists (ASA) classification.

Histopathological assessment included tumor subtype, tumor grade, and the Ki-67 proliferation index. Postoperative complications were classified according to the Clavien-Dindo system.

Renal function was evaluated using perioperative serum creatinine levels. Postoperative acute kidney injury (AKI) was defined according to kidney disease: improving global outcomes (KDIGO) 2024 criteria[15].

Patients were followed with clinical and radiological assessments at regular intervals. Local recurrence was defined as radiological or histological evidence of tumor recurrence within the retroperitoneal space. Disease-free survival was defined as the interval between surgery and recurrence or death, and overall survival was defined as the interval between surgery and death from any cause.

Statistical Analysis

Statistical analysis was performed using Jamovi software (version 2.6; The Jamovi Project (2005), Sydney, Australia). Continuous variables were analyzed using appropriate parametric or non-parametric tests, depending on their distribution. Categorical variables were compared using the chi-square test or Fisher's exact test. Survival outcomes were estimated using the Kaplan-Meier method and compared using the log-rank test. A p-value below 0.05 was considered statistically significant.

Results

A total of 21 patients underwent complete resection, including 6 nephrectomies and 15 kidney-sparing cases. The median age was 51 years; 57.1% were female. Baseline characteristics, including age, sex distribution, ASA classification, and comorbidity burden, were comparable between the two groups, with no statistically significant differences (Table 1).

Median tumor size was 27 cm [interquartile range (IQR) 21]; there was no significant difference in tumor size between groups. Adjacent organ invasion occurred in 33.3% of participants in both cohorts. DDLPS liposarcoma predominated, especially in the nephrectomy group, and exhibited higher Ki-67 values. Well-differentiated tumors were more common in kidney-sparing cases, and no R2 resections were observed (Table 2).

Preoperative creatinine was similar between groups (median 0.63 mg/dL vs. 0.78 mg/dL, p=0.21); the postoperative increase after nephrectomy was mild (median 0.95 mg/dL), clinically insignificant, and not statistically different from that after kidney-sparing surgery (median 0.72 mg/

Table 1. Demographic data and baseline clinical characteristics

Parameter	All patients (n=21)	Kidney-sparing (n=15)	Nephrectomy (n=6)	p-value
Age at diagnosis, year (IQR)	51 (17)	54 (20)	50.5 (5.5)	0.61
Female sex, n (%)	12 (57.1%)	7 (46.7%)	5 (83.3%)	0.17
ASA score, n (%)				0.69
ASA I	7 (33.3%)	6 (40.0%)	1 (16.7%)	
ASA II	10 (47.6%)	6 (40.0%)	4 (66.7%)	
ASA III	4 (19.0%)	3 (20.0%)	1 (16.7%)	
CCI score (IQR)	3 (2)	3 (2)	3 (1.5)	0.77

IQR: Interquartile range, ASA: American Society of Anesthesiologists, CCI: Charlson comorbidity index

dL, p=0.18). Postoperative outcomes were comparable, with a similar length of hospital stay between the groups. Major complications occurred only in the kidney-sparing group. No patient developed AKI according to the KDIGO 2024 criteria, and none required renal replacement therapy, indicating preserved renal function after radical nephrectomy.

After a median follow-up of 44 months, oncological outcomes favored the nephrectomy group despite a higher prevalence of aggressive histology. Local recurrence occurred in 50.0% of patients in the nephrectomy group, compared with 73.3% of patients in the kidney-sparing group (p=0.35). Median disease-free survival was longer in the nephrectomy cohort (43.5 months, IQR: 32) than in the kidney-sparing group (31 months, IQR: 43), although the difference did not reach statistical significance (p=0.29).

Long-term survival outcomes demonstrated a clinically meaningful divergence. The 5-year disease-free survival rate was 66.7% in the nephrectomy group compared to 25.0% in the kidney-sparing group. Similarly, the 5-year overall survival rate was 100% for patients who underwent nephrectomy compared with 62.0% for those treated with kidney preservation (p=0.73). A higher proportion of patients in the nephrectomy cohort received adjuvant chemotherapy (66.7% vs. 40.0%, p=0.36), reflecting preserved postoperative physiological status (Table 3).

Discussion

The surgical management of RPLS embodies one of the most complex paradigms in modern surgical oncology. Balancing the imperative to secure wide microscopic margins against

Table 2. Histopathological characteristics of the tumors

Parameter	All patients (n=21)	Kidney-sparing (n=15)	Nephrectomy (n=6)	p-value
Tumor size, cm (IQR)	27 (21)	30 (20)	22 (13)	0.17
Invasion to other organs, n (%)	7 (33.3%)	5 (33.3%)	2 (33.3%)	1.00
Tumor histology, n (%)				0.51
Well-differentiated	6 (28.6%)	5 (33.3%)	1 (16.7%)	
Myxoid type	3 (14.3%)	3 (20.0%)	0 (0.0%)	
Dedifferentiated	12 (57.1%)	7 (46.7%)	5 (83.3%)	
Tumor grade, n (%)				0.14
Grade I	1 (6.7%)	0 (0.0%)	1 (20.0%)	
Grade II	6 (40.0%)	3 (30.0%)	3 (60.0%)	
Grade III	8 (53.3%)	7 (70.0%)	1 (20.0%)	
Ki-67 index, % (IQR)	13.5 (15)	12 (12)	22 (28)	0.17
Positive margin (R2), n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	N/A

IQR: Interquartile range, N/A: Not applicable

Table 3. Postoperative morbidity and long-term oncological survival

Parameter	All patients (n=21)	Kidney-sparing (n=15)	Nephrectomy (n=6)	p-value
Length of stay, days (IQR)	6 (6)	6 (5)	6.5 (5.5)	0.78
Clavien-Dindo ≥3, n (%)	2 (9.5%)	2 (13.3%)	0 (0.0%)	1.00
KDIGO 2024 AKI incidence, n (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)	N/A
Received adjuvant chemotherapy, n (%)	10 (47.6%)	6 (40.0%)	4 (66.7%)	0.36
Local recurrence, n (%)	14 (66.7%)	11 (73.3%)	3 (50.0%)	0.35
Disease-free survival, months (IQR)	36 (44)	31 (43)	43.5 (32)	0.29
5-year DFS rate, %	36.0%	25.0%	66.7%	-
Overall survival, months (IQR)	62 (44)	77 (68)	61 (17)	0.73
5-year OS rate, %	72.0%	62.0%	100.0%	-

IQR: Interquartile range, N/A: Not applicable, KDIGO: Kidney disease: improving global outcomes, AKI: Acute kidney injury, DFS: Disease-free survival, OS: Overall survival

the morbidity of multivisceral organ sacrifice requires exceptional clinical judgment. The primary aim of this study was to critically analyze the impact of en bloc radical nephrectomy on local control, survival, and renal function, which was interpreted through the lens of modern KDIGO 2024 AKI criteria[15]. Our findings yield crucial insights that challenge the growing trend of absolute organ preservation and demonstrate that aggressive anatomical clearance can overcome malignant biological behavior without compromising patient physiology.

The retroperitoneum lacks the rigid fascial boundaries found in the extremities, facilitating microscopic tumor infiltration along the perirenal fat and Gerota's fascia[9]. While the updated Transatlantic Australasian Retroperitoneal Sarcoma Working Group guidelines advocate for an individualized, histology-driven approach to organ preservation, emphasizing the avoidance of unnecessary nephrectomies in indolent well-differentiated liposarcoma, the presence of DDLPS elements shifts the biological reality[12]. DDLPS variants exhibit ferocious local invasiveness and a high propensity for multifocal recurrence[6].

Our cohort unveiled a remarkable "histological paradox". The nephrectomy cohort was biologically disadvantaged on every measurable metric; notably 83.3% of these tumors harbored DDLPS histology and had a median Ki-67 proliferation index of 22%. According to standard tumor biology, this group should have experienced rapid local failure and high mortality. Paradoxically, radical nephrectomy reversed this outcome. The local recurrence rate was reduced by an absolute 23.3% (50.0% vs. 73.3%), and the median disease-free survival was extended by more than a year (43.5 vs. 31.0 months). Notably, the 5-year overall survival reached 100% in the nephrectomy cohort, effectively neutralizing the lethal threat posed by DDLPS histology. This confirms that visually apparent "safe" avascular planes between the tumor pseudocapsule and the kidney are often microscopically compromised in DDLPS. Removal of the ipsilateral kidney provides a definitive, inviolable anatomical margin that dictates long-term survival.

The primary argument against en bloc nephrectomy is that it induces chronic kidney disease, which significantly impairs the patient's quality of life and limits future therapeutic options. Recent massive analyses from the USSC database demonstrated that nephrectomy in RPS patients dramatically increases the risk of postoperative AKI (OR 5.16) and acute renal failure[13].

In stark contrast, our study demonstrated an absolute 0% incidence of postoperative AKI utilizing the highly sensitive KDIGO 2024 criteria[18]. This exceptional renal resilience is primarily attributed to compensatory renal hyperfiltration. In patients meticulously selected to have a healthy, structurally normal contralateral kidney and an optimal baseline creatinine (median 0.63 mg/dL in our nephrectomy group), the solitary kidney rapidly upregulates its glomerular filtration rate to accommodate the functional loss.

This preservation of renal function holds immense strategic value in the current era of sarcoma management. The paradigm of RPS treatment is currently being redefined by the international EORTC STRASS-2 phase III clinical trial, which investigates the efficacy of histology-tailored neoadjuvant chemotherapy (doxorubicin and ifosfamide) for high-risk DDLPS and leiomyosarcoma[14]. Ifosfamide is well known to be nephrotoxic, causing direct tubular damage. If a patient requires salvage chemotherapy for a future recurrence, a compromised glomerular filtration rate resulting from an injudicious prior nephrectomy will permanently disqualify them from receiving these potentially curative agents. Our data reassure sarcoma surgeons that, in the absence of preexisting nephropathy or severe diabetes, unilateral en bloc nephrectomy does not precipitate AKI and safely preserves physiological reserve, thereby allowing future nephrotoxic systemic therapies such as those utilized in STRASS-2 protocols.

Study Limitations

The primary limitation of this study is its retrospective, single-center design, which inherently introduces selection bias. Surgeons logically opt for radical nephrectomy in larger, more anatomically complex, and biologically aggressive tumors, which skews the baseline histology of the cohorts. Furthermore, the limited sample size (n=21) must be addressed mathematically rather than treated simply as a flaw. Sample size constraints resulted in a classic type II statistical error. For instance, the absolute 38% difference in 5-year OS (100% after nephrectomy vs. 62.0% after kidney-sparing) represents a substantial clinical benefit. However, the p-value for this difference in survival (p=0.73) failed to reach the conventional threshold of significance due solely to insufficient statistical power, and not to a lack of clinical efficacy. Therefore, the profound clinical benefits, in terms of local control and survival, conferred by nephrectomy should not be dismissed by a strict binary interpretation of p-values.

Conclusion

En bloc radical nephrectomy may improve local control and survival outcomes in patients undergoing resection for RPLS, particularly in cases with aggressive tumor characteristics. Importantly, this approach does not appear to compromise renal function when applied in appropriately selected patients. A tailored surgical strategy that prioritizes oncological clearance over routine organ preservation should be considered in high-risk cases.

Ethics

Ethics Committee Approval: Institutional ethical approval was obtained from Ege University prior to data collection (approval no: 24-12.1T/62, date: 26.12.2024).

Informed Consent: Retrospective observational study.

Footnotes

Paperpal and ChatGPT were used for language editing to improve grammar, readability, and clarity.

Authorship Contributions

Surgical and Medical Practices: B.G., T.G., A.U., C.Ç., O.B., Concept: B.G., K.E., Design: B.G., K.E., A.U., C.Ç., O.B., Data Collection or Processing: E.K., S.V., T.G., Analysis or Interpretation: K.E., S.V., Literature Search: E.K., T.G., Writing: B.G., K.E., E.K.

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

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

References

1. Vijay A, Ram L. Retroperitoneal liposarcoma: a comprehensive review. Am J Clin Oncol. 2015;38:213-9.
2. Jonczak E, Grossman J, Alessandrino F, Seldon Taswell C, Velez-Torres JM, Trent J. Liposarcoma: a journey into a rare tumor's epidemiology, diagnosis, pathophysiology, and limitations of current therapies. Cancers (Basel). 2024;16:3858.
3. Sbaraglia M, Bellan E, Dei Tos AP. The 2020 WHO classification of soft tissue tumours: news and perspectives. Pathologica. 2021;113:70-84.
4. Bill KL, Casadei L, Prudner BC, Iwenofu H, Strohecker AM, Pollock RE. Liposarcoma: molecular targets and therapeutic implications. Cell Mol Life Sci. 2016;73:3711-8.
5. Istl AC, Gronchi A. Neoadjuvant therapy for primary resectable retroperitoneal sarcomas-looking forward. Cancers (Basel). 2022;14:1831.
6. Gronchi A, Miceli R, Shurell E, et al. Outcome prediction in primary resected retroperitoneal soft tissue sarcoma: histology-specific overall survival and disease-free survival nomograms built on major sarcoma center data sets. J Clin Oncol. 2013;31:1649-55.
7. Anaya DA, Lev DC, Pollock RE. The role of surgical margin status in retroperitoneal sarcoma. J Surg Oncol. 2008;98:607-10.
8. Kirane A, Crago AM. The importance of surgical margins in retroperitoneal sarcoma. J Surg Oncol. 2016;113:270-6.
9. Raut CP, Swallow CJ. Are radical compartmental resections for retroperitoneal sarcomas justified? Ann Surg Oncol. 2010;17:1481-4.
10. Bonvalot S, Rivoire M, Castaing M, et al. Primary retroperitoneal sarcomas: a multivariate analysis of surgical factors associated with local control. J Clin Oncol. 2009;27:31-7.
11. Gronchi A, Lo Vullo S, Fiore M, et al. Aggressive surgical policies in a retrospectively reviewed single-institution case series of retroperitoneal soft tissue sarcoma patients. J Clin Oncol. 2009;27:24-30.
12. Stahl CC, Schwartz PB, Ethun CG, et al. Renal function after retroperitoneal sarcoma resection with nephrectomy: a matched analysis of the united states sarcoma collaborative database. Ann Surg Oncol. 2021;28:1690-6.
13. Swallow CJ, Strauss DC, Bonvalot S, et al. Management of primary retroperitoneal sarcoma (RPS) in the adult: an updated consensus approach from the transatlantic Australasian RPS working group. Ann Surg Oncol. 2021;28:7873-88.
14. Lambdin J, Ryan C, Gregory S, et al. A randomized phase III study of neoadjuvant chemotherapy followed by surgery versus surgery alone for patients with high-risk retroperitoneal sarcoma (STRASS2). Ann Surg Oncol. 2023;30:4573-5.
15. Kidney Disease: Improving Global Outcomes (KDIGO) CKD Work Group. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. Kidney Int. 2024;105:S117-314.

DOI: 10.4274/anatoljmed.2026.14238
Anatol J Gen Med Res 2026;36(2):190-9

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Multimorbidity Profiles and Health-related Quality of Life in Children with Cerebral Palsy: A Multidimensional Analysis

Serebral Palsili Çocuklarda Multimorbidite Profilleri ve Sağlıkla İlişkili Yaşam Kalitesi: Çok Boyutlu Bir Analiz

Müge Baykan[1], Özge Baykan Çopuroğlu[2]

[1]Recep Tayyip Erdoğan University Training and Research Hospital, Department of Pediatric Neurology, Rize, Türkiye
[2]Kayseri University, İncesu Ayşe and Saffet Arslan Health Services Vocational School, Department of Therapy and Rehabilitation, Kayseri, Türkiye

Cite as: Baykan M, Baykan Çopuroğlu Ö. Multimorbidity profiles and health-related quality of life in children with cerebral palsy: a multidimensional analysis. Anatol J Gen Med Res. 2026;36(2):190-9

Abstract

Objective: To quantify the relative contributions of motor severity and non-motor comorbidities to health-related quality of life (HRQoL) in individuals with cerebral palsy (CP) and to evaluate their statistical associations.

Methods: In this cross-sectional study, 218 children with CP (aged 6-18 years) were evaluated using the gross motor function classification system (GMFCS), pediatric evaluation of disability inventory (PEDI), pediatric quality of life inventory (PedsQL), and Pittsburgh sleep quality index. Comorbidities were identified through clinical records.

Results: A higher GMFCS level was the strongest independent predictor of lower HRQoL (β=-0.58, p<0.001), followed by poorer sleep quality (β=-0.25, p=0.002), epilepsy (β=-0.22, p=0.004), and malnutrition (β=-0.18, p=0.010). Dominance analysis showed that GMFCS and sleep quality together accounted for most of the explained variance in HRQoL. Latent class analysis identified three comorbidity clusters: low-burden (43%), moderate-burden (33%), and high-burden (24%), with the high-burden cluster demonstrating markedly lower PEDI and PedsQL scores (p<0.001). Sleep quality was indirectly and significantly associated with the relationships among epilepsy, malnutrition, and HRQoL.

Conclusion: Motor severity explains a large proportion of HRQoL differences in CP; however, non-motor comorbidities-especially sleep disturbance, epilepsy, and malnutrition-represent substantial and potentially modifiable contributors. Distinct multimorbidity clusters and identified mediating pathways indicate the need for integrated care models that address motor and systemic factors simultaneously. Early targeting of sleep and nutritional health may meaningfully improve quality of life trajectories in children with CP.

Keywords: Cerebral palsy, quality of life, sleep disturbance, epilepsy, malnutrition, multimorbidity

Address for Correspondence/Yazışma Adresi: Müge Baykan, MD, Recep Tayyip Erdoğan University Training and Research Hospital, Department of Pediatric Neurology, Rize, Türkiye
E-mail: dr.mugebaykan@gmail.com
ORCID ID: orcid.org/0000-0002-8299-6926

Received/Geliş tarihi: 17.03.2026
Accepted/Kabul tarihi: 17.06.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

Amaç: Bu çalışmanın amacı, serebral palside (SP) motor şiddet ile non-motor komorbiditelerin sağlıkla ilişkili yaşam kalitesi (HRQoL) üzerindeki göreceli ve istatistiksel olarak ilişkili katkılarını nicel olarak değerlendirmektir.

Yöntem: Bu kesitsel çalışmada, yaşları 6-18 arasında değişen 218 SP'li çocuk değerlendirilmiştir. Değerlendirmelerde gross motor fonksiyon sınıflandırma sistemi (GMFCS), pediyatrik engellilik değerlendirme envanteri (PEDI), pediyatrik yaşam kalitesi ölçeği (PedsQL) ve Pittsburgh uyku kalitesi indeksi kullanılmıştır. Komorbiditeler klinik kayıtlar üzerinden belirlenmiştir.

Bulgular: Daha yüksek GMFCS düzeyi, daha düşük yaşam kalitesinin en güçlü bağımsız belirleyicisi olarak bulunmuştur (β=-0,58, p<0,001). Bunu sırasıyla kötü uyku kalitesi (β=-0,25, p=0,002), epilepsi (β=-0,22, p=0,004) ve malnütrisyon (β=-0,18, p=0,010) izlemiştir. Dominans analizi, GMFCS ve uyku kalitesinin birlikte HRQoL'de açıklanan varyansın büyük kısmını oluşturduğunu göstermiştir. Latent sınıf analizi ile üç komorbidite kümesi tanımlanmıştır: düşük yük (%43), orta yük (%33) ve yüksek yük (%24). Yüksek yük grubunda PEDI ve PedsQL skorları anlamlı derecede daha düşük bulunmuştur (p<0,001). Uyku kalitesinin, epilepsi, malnütrisyon ve HRQoL arasındaki ilişkilerle anlamlı dolaylı istatistiksel ilişkiler gösterdiği belirlenmiştir.

Sonuç: Motor şiddet, SP'de yaşam kalitesindeki farklılıkların önemli bir kısmını açıklamaktadır. Ancak uyku bozukluğu, epilepsi ve malnütrisyon gibi non-motor komorbiditeler de önemli ve potansiyel olarak değiştirilebilir belirleyicilerdir. Tanımlanan multimorbidite kümeleri ve aracılık yolları, motor ve sistemik faktörleri birlikte ele alan entegre bakım modellerinin gerekliliğini ortaya koymaktadır. Özellikle uyku ve beslenme durumuna erken müdahale, SP'li çocuklarda yaşam kalitesi seyrini anlamlı şekilde iyileştirebilir.

Anahtar Kelimeler: Serebral palsi, yaşam kalitesi, uyku bozukluğu, epilepsi, malnütrisyon, multimorbidite

Introduction

Cerebral palsy (CP) is the leading cause of lifelong motor disability in childhood, affecting 2-3 per 1000 live births worldwide[1]. Despite its definition as a non-progressive motor disorder resulting from early brain injury, CP is now recognized as a complex neurodevelopmental condition with extensive multisystem involvement[2]. Non-motor comorbidities frequently exceed motor limitations in shaping functional independence, participation, and overall quality of life (QoL), yet remain comparatively underexplored in clinical research[3].

Among these comorbidities, epilepsy, malnutrition, and sleep disturbances are highly prevalent and clinically consequential. Epilepsy affects 40-60% of children with CP, particularly in those with bilateral spastic or dyskinetic subtypes, and is strongly associated with cognitive impairment and reduced QoL[4]. Malnutrition-arising from oropharyngeal dysphagia, gastrointestinal dysfunction, and increased metabolic expenditure-affects approximately one-third of children with CP, contributing to impaired growth, immune dysfunction, suboptimal neurodevelopment, and reduced participation in therapy[5]. Sleep disturbances are similarly pervasive: they occur in up to 75% of children with CP and include sleep fragmentation, sleep-disordered breathing, and circadian dysregulation[6]. Recent evidence suggests that sleep may serve as a mediator linking neurological instability, metabolic stress, and impaired daily functioning[7-9].

Despite the high prevalence of these conditions, their interdependent mechanisms and collective contribution to QoL remain poorly characterized. Research has traditionally examined epilepsy, nutrition, and sleep in isolation, overlooking their potential clustering and shared pathways[10]. Multimorbidity frameworks-long applied in adult medicine-are now gaining traction in developmental disability research, with emerging studies revealing distinct comorbidity phenotypes in CP that align with clinical severity and functional trajectories[11,12]. Latent class analysis (LCA) offers a robust, person-centred method to identify these clusters and has begun to uncover "high-burden" subgroups characterized by pronounced neurological and systemic instability[13].

Beyond descriptive clustering, mechanistic modelling approaches such as mediation analysis can clarify how specific comorbidities interact to influence functional outcomes. Prior research has shown that pain, fatigue, or parental mental health may partially mediate the association between motor severity and health-related QoL (HRQoL) in CP[14]. However, no study to date has examined whether sleep quality may partially mediate the effects of epilepsy and malnutrition on QoL, despite a compelling biological rationale. Sleep disruption can amplify seizure susceptibility, exacerbate nutritional deficiencies, worsen behavioural dysregulation, and diminish neuroplasticity-all of which are directly relevant to CP[7-9].

The present study addresses this gap by integrating variable-centred (regression, dominance analysis) and person-centred (LCA) methods to delineate the multimorbidity structure of epilepsy, malnutrition, and sleep disturbances in children with CP. A statistical mediation framework was further tested to evaluate whether sleep quality was indirectly associated with HRQoL in the relationships between neurological and nutritional comorbidities. This multidimensional analytic framework aligns with the growing emphasis on precision rehabilitation, which seeks to tailor interventions not only to motor subtype or severity but also to the child's broader systemic and neurodevelopmental profile[15].

By characterizing comorbidity clusters and elucidating potential interaction pathways, this study aims to identify modifiable, high-impact therapeutic targets that could meaningfully improve participation, independence, and QoL in children and adolescents with CP.

Materials and Methods

Study Design and Participants

This cross-sectional observational study included children and adolescents with CP who attended the Pediatric Neurology Department at Recep Tayyip Erdoğan University Training and Research Hospital between October 2024 and October 2025. All diagnoses were confirmed by a pediatric neurologist based on contemporary definitions of CP, and motor severity was classified using the [gross motor function classification system (GMFCS) levels I-V]. Children were eligible if they were between 6 and 18 years of age, had a confirmed diagnosis of non-progressive CP, and had a primary caregiver capable of completing proxy-report questionnaires. Participants were excluded if they presented with progressive neurological conditions or an acute medical illness at assessment, or had incomplete clinical, anthropometric, or questionnaire data. Severe sensory impairments that could compromise valid proxy reporting were also grounds for exclusion.

A total of 218 participants met the eligibility criteria and were included in the analyses. The sample size exceeded the requirement identified through an a priori power analysis performed using G*Power (version 3.1). Assuming a medium effect size (f^2=0.15) for multiple linear regression with four predictors, an alpha of 0.05, and 95% power, a minimum sample size of 129 participants was required. The final sample, therefore, provided adequate power for all multivariate and mediation models. Written informed consent was obtained from all caregivers. Ethical approval was granted by the Recep Tayyip Erdoğan University Non-Interventional Clinical Research Ethics Committee (approval no: 2025/372, date: 11.09.2025), and all procedures adhered to the Declaration of Helsinki.

Measures

Functional Independence and QoL

Functional independence was assessed using the pediatric evaluation of disability inventory (PEDI), which evaluates self-care, mobility, and social function, and has been validated across GMFCS levels[16]. HRQoL was measured using the parent-proxy version of the pediatric QoL inventory (PedsQL), which is widely used in children with neurodevelopmental disabilities because of its sensitivity to physical, emotional, and psychosocial functioning[17].

Sleep Quality

Sleep quality was evaluated using the Pittsburgh sleep quality index (PSQI). Although originally developed for adults, the PSQI has demonstrated strong reliability and convergent validity in pediatric neurodisability cohorts, including children with CP[18,19]. Furthermore, several pediatric neurodevelopmental studies have supported the feasibility of proxy-reported administration of the PSQI in children with communication and cognitive impairments, particularly when multidimensional sleep assessment is required. Actigraphy- and CSHQ-based studies have further supported its construct validity in this population[20].

Nutritional Status and Comorbidities

Nutritional status was determined using anthropometric measurements and clinical assessment for malnutrition [body mass index (BMI)-for-age <-2 standard deviation or clinician-confirmed nutritional deficiency], in accordance with recent guidelines for children with severe neurological impairment[21]. Comorbidities-including epilepsy, orthopedic complications, feeding difficulty, pain, gastroesophageal reflux, behavioral problems, communication deficits, and intellectual disability-were extracted from clinical and neurophysiological records. Operational definitions and full distributions are provided in Supplementary Table 1.

Statistical Analysis

Descriptive statistics were calculated to summarize demographic and clinical characteristics. Group differences

across GMFCS levels were assessed using one-way analysis of variance for continuous variables and chi-square tests for categorical variables. Pearson correlation coefficients were used to explore bivariate associations. All preliminary analyses were conducted in SPSS (version 28.0; IBM Corp.).

A multiple linear regression model was constructed to identify independent predictors of HRQoL (PedsQL total score), including GMFCS level, sleep quality (PSQI), epilepsy, and malnutrition. Model adequacy was evaluated using adjusted R^2, standardized β coefficients, 95% confidence intervals (CIs), and [variance inflation factors (VIF) <2.0]. Relative predictor importance was quantified by dominance analysis implemented in the relaimpo package in R (version 4.3.1), using 1000 bootstrap samples.

LCA was performed in Mplus version 8.9, using maximum likelihood estimation with robust standard errors. Competing models were evaluated using Akaike information criterion, Bayesian information criterion (BIC), adjusted BIC, entropy, and the bootstrap likelihood ratio test. A three-class solution was selected as optimal.

Mediation analysis was conducted using the PROCESS macro (model 4) in SPSS with 5000 bias-corrected bootstrap samples to test whether sleep quality was statistically associated with the relationships between epilepsy and HRQoL and between malnutrition and HRQoL. Indirect effects were considered significant when the 95% CI did not include zero.

Results

A total of 218 children and adolescents with CP were included. The mean age was 11.9±4.6 years (range: 6-18), and 44% were female. According to the GMFCS, 37% of participants were classified as mild (levels I-II), 28% as moderate (level III), and 35% as severe (levels IV-V). Gestational age (mean 35.4±3.6 weeks) and birth weight (mean 2520±710 g) did not differ significantly across GMFCS groups (p>0.05). Sex distribution and perinatal history also did not differ significantly between groups. Accordingly, these variables were not included as covariates in subsequent multivariable analyses. The prevalence of comorbidities increased consistently as motor severity increased, including epilepsy, malnutrition, intellectual disability, and visual and hearing impairments (all p<0.001; Table 1).

Comorbidities were common and showed pronounced gradients across the severity spectrum. Epilepsy was diagnosed in 40.4% of participants, rising from 15.2% in mild CP to 65.4% in severe CP. Malnutrition was identified in 43.6% of the cohort, with stepwise increases by CP severity (11.3% in mild CP, 35.0% in moderate CP, and 70.5% in severe CP). Feeding difficulties (32.1%), gastroesophageal reflux (19.2%), hip displacement (24.3%), scoliosis (18.8%), and joint contractures (30.3%) also increased markedly with higher GMFCS levels (all p<0.001). Intellectual disability and sensory impairments showed the same pattern (Table 2).

Functional and QoL measures demonstrated clear differences between GMFCS groups. The mean PEDI score was 54.2±18.7, with significantly lower scores in children with severe CP compared with other groups (p<0.001). Sleep quality, assessed by the PSQI, was impaired in most participants (mean 8.7±3.5), and scores increased with motor severity (p<0.001). The mean PedsQL total score was 61.4±17.6 and showed a marked gradient across GMFCS levels: higher motor severity was associated with

Table 1. Demographic and clinical characteristics of participants

Characteristic	Total (n=218)	Mild CP (I-II, n=80)	Moderate CP (III, n=60)	Severe CP (IV-V, n=78)	p-value
Age (years, mean ± SD)	11.9±4.6	10.4±3.8	11.8±4.2	13.6±4.5	0.004*
Female sex (%)	44	46	43	41	0.82
Gestational age (weeks, mean ± SD)	35.4±3.6	36.2±2.9	35.1±3.3	34.8±4.2	0.07
Birth weight (g, mean ± SD)	2520±710	2680±680	2490±700	2380±720	0.09
Presence of epilepsy (%)	40.4	15.2	41.7	65.4	<0.001**
Malnutrition (%)	43.6	11.3	35.0	70.5	<0.001**
Visual and/or hearing impairment (%)	28.8	9.0	24.0	48.7	<0.001**
Intellectual disability ≥ mild (%)	33.9	10.5	32.5	57.7	<0.001**

Continuous variables were analyzed using one-way analysis of variance, and categorical variables were analyzed using the chi-square test. A single asterisk (*) indicates p<0.05, and a double asterisk (**) indicates p<0.001; both denote statistically significant differences

SD: Standard deviation, CP: Cerebral palsy

substantially lower HRQoL (p<0.001). Physical and emotional functioning subdomains demonstrated comparable patterns (Table 3).

Correlation analyses showed strong associations between GMFCS level and both functional independence (PEDI; r=-0.75, p<0.001) and HRQoL (PedsQL; r=-0.72, p<0.001). Poorer sleep quality was moderately correlated with higher GMFCS levels (r=0.48, p<0.001), lower PedsQL scores (r=-0.56, p<0.001), and lower PEDI scores (r=-0.39, p<0.001). Epilepsy was significantly positively correlated with PSQI (r=0.34, p<0.001) and negatively correlated with functional and HRQoL measures. Malnutrition was also associated with lower PedsQL and PEDI scores (p<0.001) (Table 4).

Multivariate regression analysis identified four independent predictors of HRQoL: GMFCS level (β=-0.58, p<0.001), PSQI score (β=-0.25, p=0.002), epilepsy (β=-0.22, p=0.004), and malnutrition (β=-0.18, p=0.010). The model demonstrated good explanatory capacity (adjusted R^2=0.62), and no multicollinearity was detected (VIF <2.0). These findings indicate that motor severity, sleep disturbance, seizure disorder, and nutritional status each contribute uniquely to HRQoL (Table 5).

LCA revealed three comorbidity clusters as the optimal solution (BIC=4615.8; entropy=0.86). The low-burden class (43%) was characterized by minimal comorbidities and predominantly mild motor impairment. The moderate-burden class (33%) demonstrated intermediate rates of epilepsy, feeding problems, and malnutrition, with most children classified as GMFCS III. The high-burden class (24%) exhibited elevated probabilities of epilepsy, malnutrition, orthopedic complications, sleep disturbance, and intellectual disability, with a predominance of severe motor impairment. HRQoL and functional outcomes differed significantly between classes, with the lowest scores observed consistently in the high-burden group (p<0.001) (Supplementary Table 2, Figure 1).

Table 2. Prevalence of comorbidities across GMFCS severity levels

Comorbidity	Total (n=218)	Mild CP (I-II, n=80)	Moderate CP (III, n=60)	Severe CP (IV-V, n=78)	p-value
Epilepsy (%)	40.4	15.2	41.7	65.4	<0.001**
Nutritional difficulties (%)	38.5	10.0	32.0	69.2	<0.001**
Malnutrition (%)	43.6	11.3	35.0	70.5	<0.001**
Dysphagia (%)	32.1	9.1	29.0	61.5	<0.001**
Gastroesophageal reflux (%)	19.2	5.6	15.0	38.5	<0.001**
Hip dislocation (%)	24.3	6.3	19.0	46.2	<0.001**
Scoliosis (%)	18.8	3.8	14.5	39.7	<0.001**
Contractures (%)	30.3	8.9	26.5	55.1	<0.001**
Visual and/or hearing impairment (%)	28.8	9.0	24.0	48.7	<0.001**
Intellectual disability ≥ mild (%)	33.9	10.5	32.5	57.7	<0.001**

Comparisons among GMFCS groups were performed using the chi-square test for categorical variables

**: p<0.001 indicates statistically significant group differences, CP: Cerebral palsy, GMFCS: Gross motor function classification system

Table 3. Functional and quality of life measures across GMFCS levels

Measure	Total (n=218)	Mild CP (I-II, n=80)	Moderate CP (III, n=60)	Severe CP (IV-V, n=78)	p-value
PEDI total score (mean ± SD)	54.2±18.7	72.1±14.2	56.3±13.9	32.9±12.8	<0.001**
PSQI total score (mean ± SD)	8.7±3.5	6.0±2.8	8.5±3.1	11.0±3.2	<0.001**
PedsQL total score (mean ± SD)	61.4±17.6	80.2±12.5	62.8±13.8	42.9±11.7	<0.001**
Physical functioning subscale	58.3±19.2	79.5±14.1	60.7±13.5	37.6±12.9	<0.001**
Emotional functioning subscale	64.8±18.1	82.1±13.7	67.2±15.3	45.9±14.5	<0.001**

Group comparisons were conducted using one-way analysis of variance followed by Bonferroni-adjusted post-hoc tests **: p<0.001 indicates statistically significant group differences. Higher PSQI scores indicate poorer sleep quality, CP: Cerebral palsy, GMFCS: Gross motor function classification system, PSQI: Pittsburgh sleep quality index, PEDI: Pediatric evaluation of disability inventory, PedsQL: Pediatric quality of life inventory, SD: Standard deviation

Table 4. Correlation coefficients between functional, quality of life, and clinical variables

Variables	PedsQL	PEDI	PSQI	Epilepsy	Malnutrition	GMFCS
PedsQL	-	0.68**	-0.56**	-0.29**	-0.41**	-0.72**
PEDI	0.68**	-	-0.39**	-0.25**	-0.35**	-0.75**
PSQI	-0.56**	-0.39**	-	0.34**	0.28**	0.48**
Epilepsy	-0.29**	-0.25**	0.34**	-	0.22*	0.44**
Malnutrition	-0.41**	-0.35**	0.28**	0.22*	-	0.46**
GMFCS	-0.72**	-0.75**	0.48**	0.44**	0.46**	-

All correlations were calculated using Pearson's correlation coefficients

*: p<0.05, and **: p<0.001 indicate statistically significant relationships, GMFCS: Gross motor function classification system, PSQI: Pittsburgh sleep quality index, PEDI: Pediatric evaluation of disability inventory, PedsQL: Pediatric quality of life inventory

Table 5. Multivariate linear regression predicting health-related quality of life (PedsQL total score)

Predictor	Standardized β	95% confidence interval	p-value
GMFCS level	-0.58	-0.64 to -0.47	<0.001***
PSQI total score	-0.25	-0.39 to -0.12	0.002**
Epilepsy (yes)	-0.22	-0.33 to -0.09	0.004**
Malnutrition (yes)	-0.18	-0.29 to -0.06	0.010*

Model statistics, adjusted R^2=0.62, F (4, 213) =47.8, p<0.001; no multicollinearity detected (variance inflation factor <2.0)

*: p<0.05, **: p<0.01, ***: p<0.001 indicate statistical significance, GMFCS: Gross motor function classification system, PSQI: Pittsburgh sleep quality index, PedsQL: Pediatric quality of life inventory

Mediation analyses indicated that sleep quality was indirectly and significantly associated with the relationships between epilepsy and HRQoL and between malnutrition and HRQoL. The indirect effect was significant for epilepsy (β_indirect=-0.07; 95% CI: -0.12 to -0.03) and for malnutrition (β_indirect=-0.05; 95% CI: -0.10 to -0.02). Sleep quality accounted for 28% of the total effect of epilepsy on PedsQL and 25% of the total effect of malnutrition on PedsQL.

The mediation model demonstrated acceptable overall fit indices (comparative fit index=0.97, root mean square error of approximation=0.04) (Supplementary Table 3). Figure 2 shows the standardized mediation pathways linking epilepsy and malnutrition to HRQoL through sleep quality.

Dominance analysis demonstrated that GMFCS level contributed the largest proportion of explained variance in

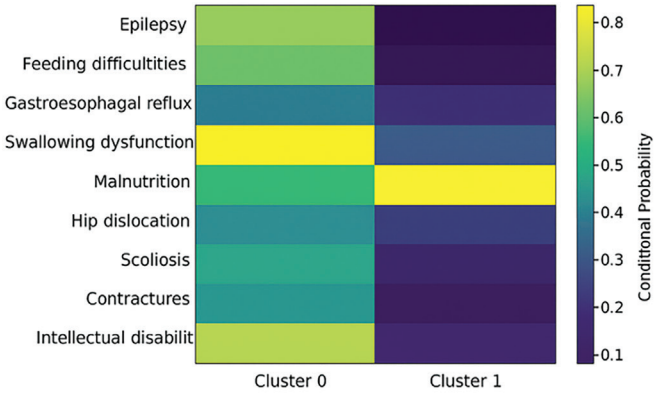


Figure 1. Latent class analysis identifying three comorbidity clusters among children with cerebral palsy. Class proportions are posterior probability-weighted estimates (43%, 33%, and 24%)

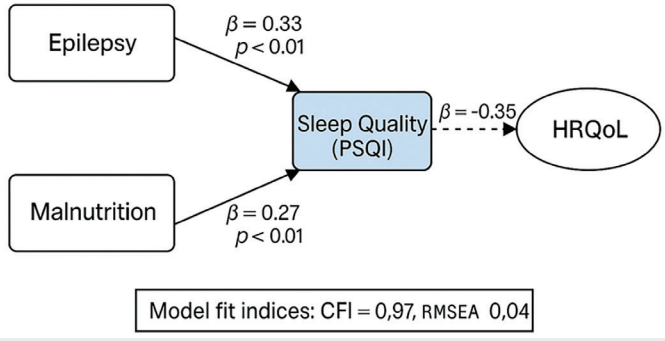


Figure 2. Standardized mediation model illustrating the statistically derived indirect associations of epilepsy and malnutrition with HRQoL through sleep quality in children with CP

CP: Cerebral palsy, PSQI: Pittsburgh sleep quality index, HRQoL: Health-related quality of life, CFI: Comparative fit index, RMSEA: Root mean square error of approximation

HRQoL (≈42%), followed by sleep quality (≈28%), epilepsy (≈18%), and malnutrition (≈12%) (Figure 3). Orthopedic variables and BMI contributed minimally.

These findings were consistent with the regression results. All variable definitions and descriptive statistics are provided in Supplementary Table 1; LCA model parameters are provided in Supplementary Table 2; and mediation coefficients are provided in Supplementary Table 3.

Discussion

This study provides a comprehensive multidimensional analysis of the determinants of HRQoL in children and adolescents with CP. The findings demonstrate that, although gross motor function remains the dominant predictor of HRQoL, sleep disturbances, epilepsy, and malnutrition exert substantial, independent effects on HRQoL. These results reinforce the contemporary understanding of CP as a multisystem neurodevelopmental condition in which functional outcomes arise from interactions among motor impairment, neurological comorbidities, and broader biopsychosocial factors[22,23].

The strong negative association between GMFCS level and HRQoL is consistent with extensive evidence linking motor severity with reduced participation, greater dependence, and limitations in daily functioning[24]. However, the regression and dominance analyses collectively indicate that motor impairment explains only part of the outcome variance.

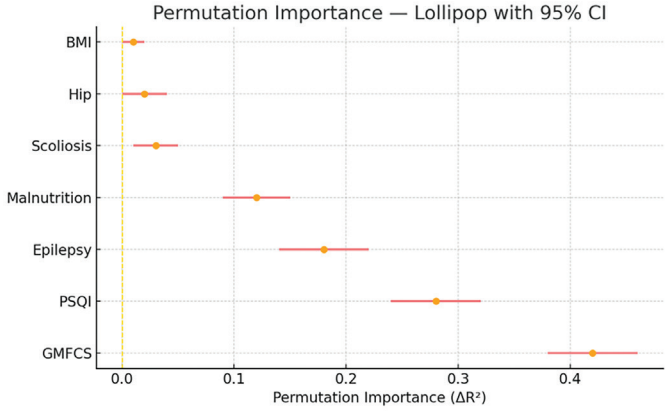


Figure 3. Relative importance (dominance weights) of predictors of health-related quality of life in children with cerebral palsy

BMI: Body mass index, CI: Confidence interval, GMFCS: Gross motor function classification system, PSQI: Pittsburgh sleep quality index

Sleep quality and nutritional status together accounted for nearly 40% of the overall explained variance, indicating that two children of the same GMFCS level may have markedly different QoL profiles depending on these modifiable factors. This supports emerging conceptual models proposing that comorbidities in CP are not merely consequences of motor dysfunction but also co-determinants of developmental, psychosocial, and functional trajectories[25].

A key contribution of the present study is the identification of sleep quality as both an independent determinant and a mediator of HRQoL. The mediation model revealed that sleep partially accounted for the associations among epilepsy, malnutrition, and HRQoL, suggesting that sleep disruption functions as an active mechanism-rather than a passive symptom-within the broader multimorbidity network in CP. Evidence from pediatric neurology increasingly demonstrates that fragmented or nonrestorative sleep is associated with cognitive impairment, behavioral dysregulation, and emotional difficulties in CP[26-28]. Longitudinal studies have shown that reduced sleep efficiency predicts poorer executive functioning and lower adaptive behavior scores over time[29]. Given its bidirectional relationship with seizures-where sleep deprivation increases epileptiform activity and nocturnal seizures degrade sleep continuity-addressing sleep disturbance may yield downstream improvements in cognitive, behavioral, and psychosocial domains[30]. Importantly, the finding that sleep quality accounted for 28% of the association between epilepsy and HRQoL suggests that therapeutic strategies for children with CP and epilepsy should extend beyond seizure control alone. Sleep-focused assessment and interventions, including sleep hygiene optimization and targeted nocturnal management, may represent important parallel components of comprehensive care in this population.

The independent contribution of malnutrition to reduced HRQoL further underscores the importance of nutritional assessment in CP. Prior research has shown that children with CP face an elevated risk of undernutrition due to dysphagia, reflux, and feeding inefficiency, but recent work also emphasizes the neurodevelopmental consequences of chronic undernutrition, including impaired cortical maturation, elevated inflammatory cytokines, and reduced neuroplastic potential[31]. The association between malnutrition and poorer sleep, observed in the mediation analysis, aligns with studies indicating shared metabolic and neuroendocrine pathways linking dietary insufficiency, disrupted circadian regulation, and increased fatigue[32]. These mechanisms offer biologically plausible explanations for the observed indirect effects on HRQoL.

The LCA expanded these findings by identifying three distinct comorbidity profiles-low-burden, moderate burden, and high burden. Children in the high-burden class exhibited markedly lower levels of HRQoL and functional independence, consistent with studies demonstrating that multimorbidity patterns in CP are structured rather than random[33,34]. Notably, the comorbidity patterns characterized here during childhood parallel those reported in adults with CP, in which neurological, musculoskeletal, and cardiometabolic multimorbidity clusters predict reduced life expectancy and increased healthcare utilization[35]. This developmental continuity suggests that early identification of high-burden profiles may enable proactive, preventive intervention to alter long-term trajectories.

Methodologically, this study contributes to the CP literature by integrating complementary statistical approaches-multivariate regression, dominance analysis, and mediation modelling-within a unified analytical framework. The regression model identified independent predictors; the dominance analysis quantified each variable's overall explanatory contribution; and the mediation analysis elucidated the statistical pathways linking comorbidities to HRQoL. Rather than conflicting, these methods converged on a coherent hierarchy: GMFCS level formed the structural basis of HRQoL, sleep quality acted as both a parallel determinant and a pathway through which epilepsy and malnutrition exert their influence, and epilepsy and malnutrition emerged as clinically meaningful comorbidities whose effects are partly modifiable. This triangulation supports a systems-level conceptualization of CP, consistent with modern models of precision rehabilitation[36,37].

The clinical implications of these findings are substantial. Routine assessment of sleep quality and nutritional status should be integrated into standard CP care pathways, alongside motor evaluation. Screening tools, such as the PSQI and structured feeding and nutrition assessments, may facilitate early detection of risk, although future work incorporating pediatric-specific sleep measures, such as the children's sleep habits questionnaire, may enhance behavioral granularity[38]. Interventions targeting sleep hygiene, nocturnal positioning, seizure optimization, and feeding rehabilitation may meaningfully improve HRQoL, potentially exceeding the effect size achievable through motor-focused therapy alone[38,39]. These results also underscore the role of interdisciplinary care models in which neurology, physiotherapy, gastroenterology, sleep medicine, and nutrition services collaborate to address the interconnected determinants of child health.

Study Limitations

Several limitations warrant consideration. The cross-sectional design precludes causal inference; it is plausible that lower HRQoL could exacerbate sleep or nutritional difficulties via reduced activity or caregiver strain. The single-center, tertiary-care sample may overrepresent children with more severe CP. A lthough PSQI demonstrates strong validity in pediatric neurodisability populations, its original development for adults should still be considered when interpreting sleep-related findings. Longitudinal, multi-center studies employing objective sleep measures, standardized nutritional assessments, and external validation of comorbidity clusters would strengthen generalizability. Although advanced statistical techniques were used, all analyses were derived from the same dataset; thus, predictive estimates require replication.

This study demonstrates that sleep quality, epilepsy, and malnutrition significantly and independently influence HRQoL in children with CP, beyond the contribution of motor severity. These findings show the importance of integrating sleep and nutritional interventions with traditional motor-based rehabilitation strategies. Recognizing CP as a multisystem condition-with modifiable comorbidities that jointly shape developmental and psychosocial outcomes-provides a strong rationale for comprehensive, interdisciplinary, and precision-oriented care models aimed at improving participation and long-term well-being for children with CP.

Conclusion

In this large cross-sectional study of children with CP, HRQoL was shaped not only by gross motor severity but also by a constellation of modifiable non-motor comorbidities. Sleep disturbance, epilepsy, and malnutrition emerged as independent and interacting determinants of well-being, with sleep quality acting as a potential explanatory factor linking neurological and nutritional burden to reduced daily functioning. The identification of distinct multimorbidity clusters and the quantification of the relative importance of predictors suggest that the cumulative impact of comorbidities is structured rather than incidental.

These findings underscore the need to reframe CP as a multisystem condition in which early recognition and management of sleep dysfunction, nutritional deficits,

and seizure instability are essential components of care. Incorporating systematic screening for these factors into routine clinical practice may meaningfully improve participation, emotional health, and long-term developmental trajectories. Future longitudinal and interventional studies are warranted to determine whether targeted, multidisciplinary approaches can modify life-course risks associated with high-burden comorbidity profiles.

Ethics

Ethics Committee Approval: Ethical approval was granted by the Recep Tayyip Erdoğan University Non-Interventional Clinical Research Ethics Committee (approval no: 2025/372, date: 11.09.2025), and all procedures adhered to the Declaration of Helsinki.

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

Acknowledgements

The authors thank the participating families and the clinical staff of the pediatric neurology clinic for their cooperation and support.

Footnotes

Authorship Contributions

Surgical and Medical Practices: M.B., Concept: M.B., Ö.B.Ç., Design: M.B., Ö.B.Ç., Data Collection or Processing: M.B., Analysis or Interpretation: M.B., Ö.B.Ç., Literature Search: M.B., Ö.B.Ç., Writing: M.B., Ö.B.Ç.

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

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

Supplementary Tables: https://d2v96fxpocvxx.cloudfront.net/df1591lb-c92e-4fb1-a3c6-b4a159ab0d2e/content-images/64bdf50c-e6b9-4913-b8ed-daa484df1c0e.pdf

References

1. Oskoui M, Coutinho F, Dykeman J, Jetté N, Pringsheim T. An update on the prevalence of cerebral palsy: a systematic review and meta-analysis. Dev Med Child Neurol. 2013;55:509-19.
2. Rosenbaum P, Paneth N, Leviton A, et al. A report: the definition and classification of cerebral palsy April 2006. Dev Med Child Neurol Suppl. 2007;109:8-14.
3. Novak I, Morgan C, Fahey M, et al. State of the evidence traffic lights 2019: systematic review of interventions for preventing and treating children with cerebral palsy. Curr Neurol Neurosci Rep. 2020;20:3.
4. Jonsson M, Hellström-Westas L, Wikman P, Ågren J. Adverse long-term outcomes following seizures in non-asphyxiated term infants: a population-based cohort study. Pediatr Res. 2026;99:112-8.
5. Calderone A, Militi D, Cardile D, Corallo F, Calabrò RS, Militi A. Swallowing disorders in cerebral palsy: a systematic review of oropharyngeal dysphagia, nutritional impact, and health risks. Ital J Pediatr. 2025;51:1-47.
6. Dos Santos Rufino A, Påhlman M, Olsson I, Himmelmann K. Characteristics and challenges of epilepsy in children with cerebral palsy-a population-based study. J Clin Med. 2023;12:346.
7. Huang X, Zhang Y, Lin Q, et al. The impact of sleep disorders on quality of life in patients with epilepsy. Seizure. 2025;129:115-22.
8. Bruni O, Breda M, Mammarella V, Mogavero MP, Ferri R. Sleep and circadian disturbances in children with neurodevelopmental disorders. Nat Rev Neurol. 2025;21:103-20.
9. Wang T, Li W, Deng J, Zhang Q, Liu Y, Zheng H. The impact of the physical activity intervention on sleep in children and adolescents with neurodevelopmental disorders: a systematic review and meta-analysis. Front Neurol. 2024;15:1438786.
10. McIntyre S, Goldsmith S, Webb A, et al. Global prevalence of cerebral palsy: a systematic analysis. Dev Med Child Neurol. 2022;64:1494-506.
11. Cooper M, Antolovich G. Comorbidities in children and adults with cerebral palsy. In: Cerebral palsy: from childhood to adulthood. Cham: Springer Nature Switzerland. 2025.
12. Leis AM, Hurvitz EA, Whitney DG. Comorbidity clusters among adults with cerebral palsy: a latent class analysis. Am J Prev Med. 2024;66:971-9.
13. Lenti MV, Pala D, Aronico N, et al. Latent class analysis identifies novel coeliac disease subgroups with distinctive clinical features: a multicentric study. Eur J Intern Med. 2025;139:106355.
14. Albayrak I, Biber A, Çalışkan A, Levendoglu F. Assessment of pain, care burden, depression level, sleep quality, fatigue and quality of life in the mothers of children with cerebral palsy. J Child Health Care. 2019;23:483-94.
15. Jackman M, Sakzewski L, Morgan C, et al. Interventions to improve physical function for children and young people with cerebral palsy: international clinical practice guideline. Dev Med Child Neurol. 2022;64:536-49.
16. Haley SM, Coster WJ, Ludlow LH, Haltiwanger JT, Andrellos PJ. Pediatric evaluation of disability inventory (PEDI): development, standardization, and administration manual. Boston, MA: New England Medical Center Hospital, PEDI Research Group; 1992.
17. Varni JW, Seid M, Rode CA. The PedsQL: measurement model for the pediatric quality of life inventory. Med Care. 1999;37:126-39.
18. Sen T, Spruyt K. Pediatric sleep tools: an updated literature review. Front Psychiatry. 2020;11:317.
19. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh sleep quality index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28:193-213.
20. Perpétuo C, Fernandes M, Veríssimo M. Comparison between actigraphy records and parental reports of child's sleep. Front Pediatr. 2020;8:567390.
21. Dipasquale V, Gottrand F, Sullivan PB, Romano C. Top-ten tips for managing nutritional issues and gastrointestinal symptoms in children with neurological impairment. Ital J Pediatr. 2020;46:35.
22. Morgan C, Fetters L, Adde L, et al. Early intervention for children aged 0 to 2 years with or at high risk of cerebral palsy: international clinical practice guideline based on systematic reviews. JAMA Pediatr. 2021;175:846-58.

23. Kassa T, Tadese H, Eriku GA, Abich Y, Fentanew M. Health-related quality of life and associated factors among primary caregivers of children with cerebral palsy, in Bahir Dar and Gondar cities, Ethiopia, 2022. PLoS One. 2024;19:e0301050.
24. Larsen SM, Terjesen T, Jahnsen RB, Diseth TH, Ramstad K. Health-related quality of life in adolescents with cerebral palsy; a cross-sectional and longitudinal population-based study. Child Care Health Dev. 2023;49:373-81.
25. Jiang L, Yang W, Chen H, Song H, Zhang S. Diagnosis and therapies for patients with cerebral palsy over the past 30 years: a bibliometric analysis. Front Neurol. 2024;15:1354311.
26. Kamara D, Bernard A, Clark ELM, et al. Systematic review and meta-analysis of behavioral interventions for sleep disruption in pediatric neurodevelopmental and medical conditions. J Pediatr Psychol. 2025;50:633-57.
27. Weighall A, Horton C. Sleep problems in childhood: understanding and addressing sleep disturbances in neurodevelopmental disorders. Journal of Family and Child Health. 2025;2:348-54.
28. van Gorp M, Dallmeijer AJ, van Wely L, et al. Pain, fatigue, depressive symptoms and sleep disturbance in young adults with cerebral palsy. Disabil Rehabil. 2021;43:2164-71.
29. Sanguino H, Brunton L, Condliffe EG, et al. Sleep characteristics of children and youth with cerebral palsy. J Child Neurol. 2025;40:123-31.
30. Gerritsen A, Hulst RY, van Rijssen IM, et al. The temporal and bi-directional relationship between physical activity and sleep in ambulatory children with cerebral palsy. Disabil Rehabil. 2024;46:2821-7.
31. Oftedal S, McCormack S, Stevenson R, Benfer K, Boyd RN, Bell K. The evolution of nutrition management in children with severe neurological impairment with a focus on cerebral palsy. J Hum Nutr Diet. 2025;38:e13277.
32. Rausch RA, Miller C, Hensler A, Goetting MG, Patel DR. Sleep disturbances and disorders in children and adolescents with cerebral palsy: a narrative review. J Clin Med. 2025;14:7828.
33. Bekteshi S, Monbaliu E, McIntyre S, et al. Towards functional improvement of motor disorders associated with cerebral palsy. Lancet Neurol. 2023;22:229-43.
34. Ryan JM, Albairami F, Hamilton T, et al. Prevalence and incidence of chronic conditions among adults with cerebral palsy: a systematic review and meta-analysis. Dev Med Child Neurol. 2023;65:1174-89.
35. Whitney DG, Kamdar NS. Development of a new comorbidity index for adults with cerebral palsy and comparative assessment with common comorbidity indices. Dev Med Child Neurol. 2021;63:313-9.
36. Hidalgo-Robles Á, Merino-Andrés J, Cisse MRS, Pacheco-Molero M, León-Estrada I, Gutiérrez-Ortega M. The pathway is clear but the road remains unpaved: a scoping review of implementation of tools for early detection of cerebral palsy. Children (Basel). 2025;12:941.
37. Faccioli S, Pagliano E, Ferrari A, et al. Evidence-based management and motor rehabilitation of cerebral palsy children and adolescents: a systematic review. Front Neurol. 2023;14:1171224.
38. Fış NP, Arman A, Topuzoğlu A, et al. The validity and the reliability of the Turkish version of the children's sleep habits questionnaire. Alpha Psychiatry. 2021;11:151-60.
39. Zh Chia A, Tan YH, Yeo TH, Teoh OH, Min Ng Z. Epidemiology and risk factors for sleep disturbances in children and youth with cerebral palsy: an ICF-based approach. Sleep Med. 2022;96:93-8.

DOI: 10.4274/anatoljmed.2026.67044
Anatol J Gen Med Res 2026;36(2):200-11

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

A Human-centered Evaluation of Artificial Intelligence and a Single Pulmonologist's Responses in Pulmonary Medicine Communication

Göğüs Hastalıkları İletişiminde Yapay Zeka ve Tek Bir Göğüs Hastalıkları Hekiminin Yanıtlarının İnsan Odaklı Değerlendirmesi

Damla Serçe Unat[1,4], Ömer Selim Unat[2,5], Mutlu Onur Güçsav[1,5], Berkay Can Karakoç[3], Aysu Ayrancı[1], Ahmet Emin Erbaycu[1]

[1]İzmir Bakırçay University Faculty of Medicine, Department of Pulmonology, İzmir, Türkiye
[2]University of Health Sciences Türkiye, İzmir Dr. Suat Seren Chest Diseases and Surgery Training and Research Hospital, İzmir, Türkiye
[3]Başkent University, Institute of Science and Technology, Department of Defense Electronics and Software, Ankara, Türkiye
[4]Ege University Faculty of Medicine, Department of Basic Oncology, İzmir, Türkiye
[5]Ege University Faculty of Medicine, Department of Translational Pulmonology, İzmir, Türkiye

Cite as: Serçe Unat D, Unat ÖS, Güçsav MO, Karakoç BC, Ayrancı A, Erbaycu AE. A human-centered evaluation of artificial intelligence and a single pulmonologist's responses in pulmonary medicine communication. Anatol J Gen Med Res. 2026;36(2):200-11

Abstract

Objective: Large language models (LLMs) are increasingly used by patients and the public to access medical information. In pulmonary medicine, where information needs span a wide range of chronic and potentially serious diseases, the quality of artificial intelligence-generated communication requires systematic evaluation.

Methods: This cross-sectional comparative study evaluated responses to 37 standardized pulmonary medicine questions covering major respiratory domains. Each question was answered by four LLMs-Gemini 2.5, GPT-4, Claude Sonnet 4, and DeepSeek v3.2-and a board-certified professor of pulmonology with more than 25 years of clinical experience. Six blinded raters, including three pulmonologists and three laypersons, assessed anonymized responses using a five-point Likert scale across conceptually aligned domains: accuracy/perceived reliability, clarity, and scientific adequacy/practical usefulness. Inter-rater reliability was assessed using intraclass correlation coefficients (ICC). Comparative analyses were performed using repeated-measures one way analysis of variance with Bonferroni-adjusted pairwise comparisons and Friedman tests.

Results: Inter-rater reliability indicated slight single-measure agreement (ICC 0.09-0.12) and fair-to-moderate average-measure agreement (ICC 0.36-0.45), which is consistent with the subjective nature of communication-focused assessment. Overall, LLM-generated responses received higher communication-quality ratings than physician-generated responses across evaluation domains. Gemini 2.5 and GPT-4 achieved the most favorable scores, particularly on accuracy/perceived reliability and clarity. Differences between LLMs and the physician were more pronounced among lay raters than among expert raters. Among expert raters, physician-generated responses were rated among the highest for accuracy and scientific adequacy, although this pattern was not observed for clarity.

Conclusion: Contemporary LLMs may provide clear, reliable, and practically useful information on pulmonary medicine, particularly as perceived by lay users, when evaluated from both expert and lay perspectives. These findings support their adjunctive use in patient education and health communication and emphasize that LLMs cannot replace clinical judgment, professional accountability, or physician oversight.

Keywords: Large language model, artificial intelligence, medical informatics, health communication, patient education, pulmonary medicine

Address for Correspondence/Yazışma Adresi: Damla Serçe Unat, MD, İzmir Bakırçay University Faculty of Medicine, Department of Pulmonology, İzmir, Türkiye
E-mail: damla.serceunat@bakircay.edu.tr
ORCID ID: orcid.org/0000-0003-4743-5469

Received/Geliş tarihi: 17.05.2026
Accepted/Kabul tarihi: 26.06.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

Amaç: Büyük dil modelleri (LLM), hastalar ve toplum tarafından tıbbi bilgiye ulaşmak amacıyla giderek daha sık kullanılmaktadır. Göğüs hastalıkları alanında bilgi gereksinimi kronik ve potansiyel olarak ciddi hastalıkları kapsadığından, yapay zeka tarafından oluşturulan tıbbi iletişim içeriğinin kalitesinin sistematik olarak değerlendirilmesi önem taşımaktadır.

Yöntem: Bu kesitsel karşılaştırmalı çalışmada, temel solunum hastalıkları alanlarını kapsayan 37 standart göğüs hastalıkları sorusuna verilen yanıtlar değerlendirildi. Her soru dört LLM-Gemini 2.5, GPT-4, Claude Sonnet 4 ve DeepSeek v3.2-ile 25 yıldan fazla klinik deneyime sahip bir göğüs hastalıkları profesörü tarafından yanıtlandı. Üç göğüs hastalıkları uzmanı ve üç uzman olmayan değerlendiriciden oluşan altı kör değerlendirici, anonimleştirilmiş yanıtları beşli Likert ölçeği ile; doğruluk/algılanan güvenilirlik, açıklık ve bilimsel yeterlilik/pratik yararlılık alanlarında değerlendirdi. Değerlendiriciler arası güvenilirlik sınıf içi korelasyon katsayısı (ICC) ile analiz edildi. Karşılaştırmalı analizlerde tekrarlı ölçümler tek yönlü varyans analizi, Bonferroni düzeltmeli ikili karşılaştırmalar ve Friedman testi kullanıldı.

Bulgular: Değerlendiriciler arası güvenilirlik, tekil ölçümlerde zayıf (ICC 0,09-0,12), ortalama ölçümlerde ise düşük-orta ile orta düzeyde (ICC 0,36-0,45) bulundu ve bu durum iletişim odaklı değerlendirmenin öznel yapısı ile uyumluydu. Genel olarak, büyük dil modeli yanıtları tüm değerlendirme alanlarında hekim yanıtlarına göre daha yüksek iletişim kalitesi puanları aldı. Gemini 2.5 ve GPT-4 özellikle doğruluk/algılanan güvenilirlik ve açıklık alanlarında en olumlu skorları elde etti. Büyük dil modelleri ile hekim yanıtları arasındaki farklar, uzman değerlendiricilere kıyasla uzman olmayan değerlendiricilerde daha belirgindi. Uzman değerlendiriciler arasında ise hekim tarafından oluşturulan yanıtlar doğruluk ve bilimsel yeterlilik açısından en yüksek puan alan yanıtlar arasında yer alırken, bu örüntü açıklık alanında gözlenmedi.

Sonuç: Güncel büyük dil modelleri, hem uzman hem de uzman olmayan değerlendiriciler açısından açık, güvenilir ve pratik olarak yararlı göğüs hastalıkları bilgisi sunabilir. Bu bulgular, büyük dil modellerinin hasta eğitimi ve sağlık iletişiminde yardımcı araçlar olarak kullanılabileceğini desteklemektedir. Bununla birlikte, bu sistemler klinik yargının, mesleki sorumluluğun ve hekim gözetiminin yerini alamaz.

Anahtar Kelimeler: Büyük dil modelleri, yapay zeka, tıbbi bilişim, sağlık iletişimi, hasta eğitimi, göğüs hastalıkları

Introduction

Effective medical communication is a core component of high-quality healthcare delivery, since the accuracy, reliability, and clarity of the information shared with patients and caregivers directly shape clinical understanding and decision-making. Prior work has consistently demonstrated that clear and accurate responses to patient inquiries are associated with greater patient satisfaction, improved treatment adherence, and a stronger physician-patient relationship, making comprehensible and trustworthy communication central to patient-centered care[1,2]. This is particularly consequential in populations with limited health literacy, where access to accurate information from trustworthy sources is essential for both treatment success and patient safety[3].

In parallel, the integration of the internet and digital media into everyday life has reshaped how patients and their relatives seek health information, with online searches for symptoms and diagnoses giving rise to the so-called "Dr. Google" phenomenon[4]. Much of the health-related material available online, however, is inaccurate, incomplete, or contextually misleading[5], with the potential to provoke unnecessary anxiety, misdirected care-seeking, or distrust toward medical advice. From a public health perspective, the dissemination of scientifically accurate, comprehensible, and trustworthy health information through digital channels has therefore become both an individual and a collective responsibility[6].

Within this evolving information ecosystem, artificial intelligence (AI) has progressively moved from a peripheral analytic tool to a core component of contemporary medical informatics. Earlier generations of AI demonstrated value in image-based diagnostic support, bioinformatics pipelines, and the interpretation of large-scale data from genetic biobanks, where machine learning approaches enabled scalable analyses of complex, heterogeneous datasets that would be impractical to perform manually[7-9]. Building on this foundation, AI systems have since been embedded into clinical decision-support workflows, risk stratification models, and individualized treatment planning informed by genomic, proteomic, and radiomic data[10,11]. What distinguishes the current phase, however, is the migration of AI from structured data and pixel-level analysis toward the communication layer of medicine-the space in which patients, caregivers, and clinicians exchange information, weigh options, and arrive at decisions. This shift reframes AI not only as an analytic engine but also as an active participant in health information delivery and brings the quality, transparency, and safety of AI-generated communication into the foreground of medical informatics research.

Beyond diagnostics and data analytics, large language models (LLMs) are transforming the communication layer of medicine. Leveraging advanced natural language processing, these systems have shown promise in medical question answering, summarization of clinical dialogues, and analysis of unstructured text from electronic health records[11,12]. Their use in healthcare has expanded rapidly, with domain-specific implementations developed to improve the accuracy, clarity, and empathy of responses to medical queries[11]. At the same time, their deployment raises non-trivial concerns regarding factual reliability, contextual nuance, data security, algorithmic bias, and the risk of confidently incorrect outputs-issues that underscore the need for safe, transparent, and human-centered integration of AI technologies into clinical practice[7]. Systematic evaluation of LLM-generated responses, especially within clinically meaningful, domain-specific contexts such as pulmonology, is therefore essential before such tools can be responsibly recommended for patient-facing use[13].

These considerations are particularly salient in pulmonary medicine. Respiratory conditions-including asthma, chronic obstructive pulmonary disease (COPD), pneumonia, pulmonary thromboembolism, interstitial lung diseases, obstructive sleep apnea, tuberculosis, and lung cancer-account for a substantial share of the global disease burden and result in ongoing public information needs. Many of these conditions are chronic, require sustained self-management, and involve decisions regarding inhaler use, oxygen therapy, smoking cessation, vaccination, surgical options, and end-of-life care. The accessibility, accuracy, and clarity of the information delivered to patients and their relatives therefore have direct implications for adherence, safety, and informed decision-making. Despite the rapid uptake of LLMs in this space, the available evidence remains fragmented: many existing evaluations examine a single model, rely exclusively on expert reviewers, or assess technical accuracy without considering how lay audiences perceive the information. Studies comparing multiple state-of-the-art LLMs head-to-head with an experienced physician on the same set of publicly relevant questions-using blinded evaluations by both clinicians and laypersons across complementary criteria of accuracy, clarity, and scientific or practical adequacy-remain limited. This gap is consequential because expert and lay perspectives often diverge: clinicians tend to weigh factual completeness and technical precision, whereas patients and the broader public appraise responses through the lens of perceived reliability, comprehensibility, and real-world usefulness. Capturing both perspectives is therefore essential to characterize the realistic role of LLMs in patient-facing health communication and decision-support tasks.

The present study was designed to address this gap. We conducted a comparative, blinded evaluation in which four contemporary LLMs (Gemini 2.5, GPT-4, Claude Sonnet 4, and DeepSeek v3.2) and a board-certified pulmonology professor with more than 25 years of clinical experience responded to the same set of 37 standardized pulmonary medicine questions derived from publicly observable information-seeking behavior. Using conceptually aligned criteria (accuracy/perceived reliability, clarity, and scientific adequacy/practical usefulness), anonymized responses were independently rated by three expert pulmonologists and three laypersons. By examining inter-rater reliability and systematically comparing model and physician performance across audience types, the study provides an informatics-oriented assessment of how well current LLMs support communication about pulmonary medicine to non-specialist users, and aims to clarify their potential-and limitations-as complementary tools alongside experienced clinicians in health information delivery and clinical decision-support workflows.

Materials and Methods

This study was a cross-sectional, comparative evaluation designed to assess the quality and reliability of responses generated by four LLMs and a human pulmonology expert.

Design and Procedure

This study did not involve patients, patient data, identifiable personal information, biological samples, or any clinical intervention. The raters evaluated anonymized written responses; therefore, ethics committee approval was not required according to institutional and national regulations. Informed consent is not applicable, as the study did not involve patients or identifiable personal data.

Question Selection

A total of 37 standardized questions related to chest diseases were developed by a pulmonology specialist who had no conflicts of interest with either the expert or layperson raters. The questions were designed to represent a balanced distribution across major subspecialties within pulmonary medicine. Particular attention was paid to ensuring that the set encompassed diverse, clinically relevant domains: asthma (3), COPD (4), pneumonia (4), pulmonary thromboembolism

(4), interstitial lung diseases (3), obstructive sleep apnea (4), lung cancer (7), tuberculosis (2), and other respiratory topics (6). In addition to its clinical relevance, the study was designed to reflect real-world patterns of health information consumption, focusing on publicly accessible questions and communication formats that influence health-related understanding beyond clinical settings. The complete 37-item question set is provided as supplementary material; formal psychometric content validation (e.g., a content-validity index) was not performed.

The questions' content was informed by Google Trends data, social media discussions, and frequently asked questions directed to a pulmonologist's professional social media account, to ensure that the questions reflected publicly relevant, real-world, and digitally mediated health concerns. This approach provided a representative sampling of topics commonly encountered in both clinical practice and public health communication contexts.

Participants and Evaluation Procedure

Six independent raters participated in the evaluation: three expert pulmonologists (E1, E2, and E3, with 7, 8, and 10 years of clinical experience, respectively) and three laypersons (L1, L2, and L3, a 32-year-old university-educated teacher, a 53-year-old employee and former amateur footballer, and a 45-year-old homemaker with a high-school education). All raters were blinded to the identity of the content source (LLM-generated vs. professor-written).

Four LLMs (Gemini 2.5, GPT-4, Claude Sonnet 4, and DeepSeek v3.2) were selected because they represent the most advanced publicly available generative models at the time of the study (prompts submitted in August 2025), encompassing diverse architectures, training paradigms, and user interfaces. Their inclusion enabled a balanced assessment across distinct LLM frameworks. A pulmonology professor with more than 25 years of clinical experience provided physician-generated responses. The physician is actively engaged in clinical practice, holds an administrative role, and is involved in faculty development and educator training activities. Because a single expert served as the human comparator, this benchmark reflects the performance of one highly experienced pulmonologist rather than that of physicians in general.

Each of the six raters evaluated the anonymized answers of the five respondents (four LLMs and one human professor). Ratings were assigned on a five-point Likert scale (1=lowest, 5=highest) based on the designated evaluation criteria for each group. None of the raters had any conflict of interest with either the professor or the LLM selection process. All evaluations were performed independently, and raters were blinded to the identity of each response source to ensure unbiased judgment based solely on content quality.

Conceptual Alignment of Evaluation Criteria

The three criteria used by layperson raters-perceived reliability, clarity, and practical usefulness-were designed to conceptually parallel the expert criteria of accuracy, clarity, and scientific adequacy, respectively. Perceived reliability corresponds to accuracy, as both reflect the extent to which information is regarded as correct and trustworthy, albeit from different perspectives: experts assess factual correctness, whereas laypersons judge credibility and the confidence it inspires. Clarity was retained consistently across groups, reflecting the mutual importance of linguistic simplicity and structural transparency. Finally, practical usefulness parallels scientific adequacy: experts evaluate the scientific rigor and completeness of the answer, whereas laypersons perceive the same quality in terms of real-world applicability and utility.

The criteria used for layperson evaluations were adapted from the methodological framework applied in the study by Charide et al.[14], which focused on public and user-centered assessment approaches in health communication research. In parallel, the expert evaluation criteria were determined following the methodological approach used by Akçay et al.[15] in their study evaluating the accuracy and reliability of ChatGPT 4's responses in thoracic surgery. This mapping ensures conceptual equivalence across evaluator groups while maintaining relevance to their respective levels of expertise. Because expert accuracy (factual correctness) and lay perceived reliability are related but non-identical constructs, inter-rater reliability is reported separately for experts and laypersons, and the pooled coefficient is interpreted with caution.

Procedural Neutrality and Data Collection

The questions were submitted to the LLMs by an independent engineer specializing in software development, who is one of the authors of this article. This engineer had no background in medicine and no prior acquaintance or conflict of interest with the participating pulmonology professor. Each LLM was reset to a new session before being questioned to prevent memory retention or contextual bias, ensuring that each

interaction represented a clean, independent exchange. All prompts were entered using the same computer, under identical network conditions and internet speed, to standardize response latency and interface performance. The engineer took no part in the writing, scoring, or evaluation stages and did not participate as either an expert or layperson rater, thereby ensuring full procedural neutrality and minimizing potential bias throughout data collection.

Statistical Analysis

Inter-rater reliability was assessed using intraclass correlation coefficients (ICC) to quantify agreement among raters. ICC values estimate the proportion of variance in ratings that is attributable to true score differences rather than to random error or rater subjectivity. Both single-measure and average-measure ICCs (two-way random effects model, absolute agreement type) were computed to capture individual and group-level reliability[16].

Interpretation of ICC values followed established benchmarks, with values below 0.20 indicating slight agreement, 0.21-0.40 fair agreement, 0.41-0.60 moderate agreement, 0.61-0.80 substantial agreement, and 0.81-1.00 almost perfect agreement[17].

Comparative analyses across the five response sources were performed using repeated-measures one way analysis of variance (ANOVA) because each rater provided scores for all sources, resulting in a within-subjects (dependent) data structure. This approach allowed comparison of mean ratings across conditions while accounting for the interdependence of repeated observations from the same raters. Bonferroni-adjusted pairwise comparisons were applied to control for type I error in multiple testing. To corroborate these findings and to address potential deviations from normality, Friedman tests were additionally conducted as non-parametric equivalents. The unit of analysis was the per-question mean across raters (n=37), with response source as the within-subject factor (degrees of freedom=4, 144). Sphericity was assessed (Greenhouse-Geisser), and a linear mixed-effects model with crossed random effects for questions and raters was acknowledged as a valid alternative.

To ensure robustness and to explore subgroup consistency, three complementary analytic models were applied: (1) all raters (E1-E3, L1-L3), (2) experts only (E1-E3), and (3) laypersons only (L1-L3). For each model, Bonferroni-adjusted pairwise comparisons were used to compare the five response sources across the three evaluation criteria. In these analyses, accuracy for expert raters was conceptually aligned with perceived reliability for layperson raters; clarity was shared by both groups; and scientific adequacy for expert raters was aligned with practical usefulness for layperson raters.

The primary analyses included all six raters, with subgroup analyses conducted separately for expert and layperson raters. Statistical analyses were conducted using IBM SPSS Statistics version 26 (IBM Corp., Armonk, NY, USA), with recomputation and verification performed using Python and the pingouin package.

Results

Inter-rater Reliability

Inter-rater reliability was assessed using a two-way random-effects model (absolute agreement). As summarized in Table 1, single-measure ICCs ranged from 0.09 to 0.12 and average-measure ICCs ranged from 0.36 to 0.45 across accuracy/perceived reliability, clarity, and scientific adequacy/practical usefulness (all $p<0.001$).

According to commonly applied interpretative frameworks, these values correspond to slight agreement for single measures and to fair-to-moderate agreement for average measures, indicating variability in individual ratings but improved consistency when scores were aggregated. Differences in scoring patterns across evaluators and domains were statistically significant ($p<0.001$), reflecting heterogeneity in perception rather than systematic measurement error.

Overall, these findings demonstrate that while individual-level agreement was limited, aggregated ratings showed greater consistency, underscoring the subjective nature of communication-based evaluations and the influence of evaluator background on inter-rater agreement metrics.

Summary of Evaluator Scores

Across evaluation domains, the relative ranking of mean scores remained stable, with LLM-generated responses generally receiving higher ratings than physician-generated responses. Differences in mean ratings across models and evaluator groups reflected variability in perceived quality rather than isolated or extreme scoring behavior. The observed score distributions indicate coherent, interpretable evaluation patterns across all six raters and support the internal consistency of the comparative assessment framework.

Table 1. Inter-rater reliability (intraclass correlation coefficients), by rater group

Criterion	Single ICC (95% CI)	Average ICC (95% CI)	Single	Average
All raters (n=6)				
Accuracy/perceived reliability	0.087 (0.04, 0.14)	0.364 (0.20, 0.50)	slight	fair
Clarity	0.118 (0.07, 0.18)	0.445 (0.31, 0.56)	slight	moderate
Scientific adequacy/practical usefulness	0.117 (0.07, 0.17)	0.442 (0.31, 0.56)	slight	moderate
Experts only (E1-E3)				
Accuracy/perceived reliability	0.194 (0.10, 0.29)	0.419 (0.25, 0.55)	slight	moderate
Clarity	0.161 (0.07, 0.25)	0.365 (0.20, 0.51)	slight	fair
Scientific adequacy/practical usefulness	0.207 (0.10, 0.32)	0.439 (0.25, 0.58)	slight	moderate
Laypersons only (L1-L3)				
Accuracy/perceived reliability	0.096 (0.00, 0.20)	0.243 (0.01, 0.43)	slight	fair
Clarity	0.148 (0.05, 0.25)	0.342 (0.14, 0.50)	slight	fair
Scientific adequacy/practical usefulness	0.192 (0.10, 0.29)	0.415 (0.24, 0.55)	slight	moderate

Accuracy (experts)/perceived reliability (laypersons), scientific adequacy (experts)/practical usefulness (laypersons) Two-way random-effects model, absolute agreement; 185 rated items. Bands per Landis and Koch[17], ICC: Intraclass correlation coefficients, CI: Confidence interval

When all six raters were included, all four LLMs received higher mean ratings across all evaluation criteria than did physician-generated responses. Among the LLMs, Gemini 2.5 and GPT-4 achieved the highest overall mean scores, followed by DeepSeek v3.2 and Claude Sonnet 4. Physician-generated responses consistently received lower average ratings, with the largest differences observed in the scientific adequacy domain. Measures of dispersion were modest across all models, indicating a relatively consistent pattern of scoring among evaluators (Table 2a).

Among expert raters, the single pulmonologist was rated among the highest, rather than below the models; the physician received an accuracy/perceived reliability mean (4.66), nearly identical to Gemini 2.5, which had the highest mean score (4.67), and a scientific adequacy mean (4.13) comparable to or higher than most LLMs. Clarity scores were similar across sources. Standard deviations were small, indicating consistent expert judgments (Table 2b).

Among layperson raters, differences in mean ratings across response sources were more pronounced. Across all evaluation criteria, LLM-generated responses received higher mean scores than physician-generated responses, with the largest differences observed in accuracy and clarity. Gemini 2.5 achieved the highest and most consistent mean ratings among the LLMs. In contrast, ratings assigned to physician-generated responses showed greater dispersion, reflected by higher standard deviations, indicating increased variability in evaluations by non-expert participants (Table 2c).

When mean ratings across all 37 questions were examined within each evaluation domain (accuracy and perceived reliability, clarity, and scientific adequacy or practical usefulness), LLM-generated responses generally showed higher mean scores, whereas physician-generated responses tended to cluster toward lower values. Overlapping data points indicate identical or highly similar scores assigned by different models to specific questions, reflecting comparable scoring patterns across models rather than isolated deviations. These distributions provide a visual summary of score alignment across questions and evaluation domains (Figure 1).

Comparative Analyses

A repeated-measures ANOVA demonstrated a significant main effect of response source across all rater configurations. When all six raters were included, statistically significant differences were observed among the five response sources for all three evaluation criteria ($p<0.001$). Differences were most pronounced for accuracy and perceived reliability, both of which showed the largest effect size (partial $\eta^2=0.26$). For all raters: accuracy/perceived reliability, $F(4.144)=12.84$, partial $\eta^2=0.263$; clarity, $F(4.144)=10.47$, partial $\eta^2=0.225$; scientific adequacy/practical usefulness, $F(4.144)=11.89$, partial $\eta^2=0.248$ (all $p<0.001$). Subgroup analyses for expert and layperson raters are presented in Supplementary Tables 1-3.

When analyses were restricted to expert pulmonologists, significant effects of response source remained for accuracy

Table 2a. Mean (± SD) scores by response source-all raters			
Response source	**A/PR (mean ± SD)**	**C (mean ± SD)**	**SA/PU (mean ± SD)**
Gemini 2.5	4.69±0.54	4.70±0.60	4.43±0.68
GPT-4	4.51±0.61	4.66±0.57	4.22±0.71
DeepSeek v3.2	4.38±0.87	4.48±0.75	4.18±0.80
Claude Sonnet 4	4.38±0.70	4.57±0.60	4.10±0.70
Professor (physician)	**4.24±0.97**	**4.26±0.97**	**3.93±0.92**
SD: Standard deviation, A/PR: Accuracy (experts)/perceived reliability (laypersons), C: Clarity, SA/PU: Scientific adequacy (experts)/practical usefulness (laypersons)			

Table 2b. Mean (± SD) scores by response source-experts (E1-E3)			
Response source	**A/PR (mean ± SD)**	**C (mean ± SD)**	**SA/PU (mean ± SD)**
Gemini 2.5	4.67±0.51	4.72±0.51	4.20±0.63
GPT-4	4.48±0.60	4.68±0.51	3.92±0.61
DeepSeek v3.2	4.52±0.55	4.61±0.51	4.00±0.69
Claude Sonnet 4	4.39±0.61	4.62±0.52	3.80±0.62
Professor (physician)	**4.66±0.51**	**4.59±0.58**	**4.13±0.65**
SD: Standard deviation, A/PR: Accuracy (experts)/perceived reliability (laypersons), C: Clarity, SA/PU: Scientific adequacy (experts)/practical usefulness (laypersons)			

Table 2c. Mean (± SD) scores by response source-laypersons (L1-L3)			
Response source	**A/PR (mean ± SD)**	**C (mean ± SD)**	**SA/PU (mean ± SD)**
Gemini 2.5	4.71±0.56	4.68±0.69	4.66±0.65
GPT-4	4.55±0.61	4.64±0.63	4.52±0.67
DeepSeek v3.2	4.23±1.08	4.35±0.92	4.35±0.87
Claude Sonnet 4	4.38±0.79	4.51±0.67	4.40±0.66
Professor (physician)	**3.83±1.13**	**3.93±1.16**	**3.74±1.09**
SD: Standard deviation, A/PR: Accuracy (experts)/perceived reliability (laypersons), C: Clarity, SA/PU: Scientific adequacy (experts)/practical usefulness (laypersons)			

(perceived reliability) and for scientific adequacy or practical usefulness, whereas differences in clarity were of smaller magnitude. In contrast, among layperson raters, significant differences among response sources were consistently observed for all evaluation criteria ($p<0.001$), with larger effect sizes indicating greater differentiation in mean ratings.

Across subgroup analyses, the direction and magnitude of differences varied by rater expertise, with LLM advantages most evident among lay raters. Detailed ANOVA results for each configuration are provided in Supplementary Tables 1-3.

Given the significant main effects observed, Bonferroni-adjusted post-hoc tests were performed to determine which specific model comparisons accounted for these differences. Detailed pairwise statistics, including mean differences, 95% confidence intervals, t-values, degrees of freedom, Cohen's dz, and Bonferroni-adjusted p-values, are provided in Supplementary Tables 4-6.

When all raters were included, Bonferroni-adjusted pairwise comparisons showed that Gemini 2.5 and GPT-4 were rated significantly higher than the physician on all three criteria ($p\leq0.015$), whereas DeepSeek v3.2 did not differ significantly from the physician on any criterion, and Claude Sonnet 4 differed only for clarity ($p=0.010$). Among the models, Gemini 2.5 scored significantly higher than DeepSeek v3.2 and Claude Sonnet 4 on accuracy/perceived reliability and scientific adequacy/practical usefulness, and higher than GPT-4 on those two criteria, with no significant difference in clarity (Table 3a).

Among expert raters, no LLM was rated significantly higher than the physician on any criterion after Bonferroni adjustment; however, the physician was rated significantly higher than Claude Sonnet 4 on accuracy/perceived reliability ($p=0.046$) and scientific adequacy ($p=0.004$). Experts rated all sources similarly for clarity. These findings indicate that, among professional evaluators, the experienced pulmonologist was judged to be at least as accurate and scientifically adequate as the strongest LLMs (Table 3b).

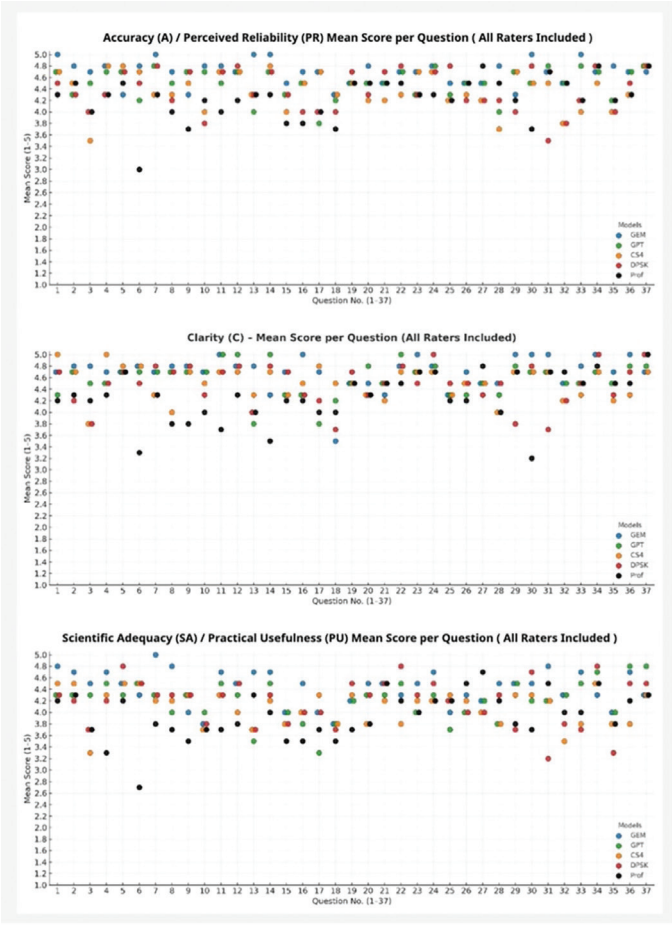


Figure 1. Distribution of mean scores per question across all raters

For each of the 37 questions, the mean Likert score (1-5, averaged across all six raters) is plotted separately for accuracy/perceived reliability, clarity, and scientific adequacy/practical usefulness. The x-axis represents the question number (1-37), and the y-axis represents the mean score. Color key: Gemini 2.5, GPT-4, Claude Sonnet 4, DeepSeek v3.2, and the physician. Overlapping markers indicate identical or closely similar means

Layperson raters' evaluations revealed statistically significant differences between response sources across all evaluation domains ($p<0.001$). Across measures of accuracy and perceived reliability, clarity, and scientific adequacy or practical usefulness, LLM-generated responses received higher mean ratings than physician-generated responses, with Gemini 2.5 and GPT-4 showing slightly higher mean ratings than Claude Sonnet 4 and DeepSeek v3.2. These differences were most pronounced for accuracy and for perceived reliability and clarity. Overall, lay evaluators consistently assigned higher ratings to LLM-generated content across domains, reflecting differences in perceived quality and usability rather than isolated rating effects (Table 3c).

Discussion

The present study demonstrated slight single-measure inter-rater reliability (ICC 0.09-0.12) and fair-to-moderate average-measure (ICC 0.36-0.45) inter-rater reliability, which is acceptable within the context of subjective, communication-focused assessments. Across evaluation domains, LLM responses received higher mean ratings than the physician did overall; however, this pattern was driven by lay raters. Expert raters rated the physician among the highest for accuracy and scientific adequacy, and no LLM significantly exceeded the physician's performance. Answer length was also positively associated with ratings across sources (Pearson $r\approx0.73$), identifying response length/format as a potential confounder. Mean response lengths differed across sources, with Gemini 2.5 producing the longest responses (approximately 33 words), followed by DeepSeek v3.2 (approximately 21 words), GPT-4 (approximately 20 words), the physician-generated responses (approximately 18 words), and Claude Sonnet 4 (approximately 14 words). Repeated-measures ANOVA confirmed a significant main effect of response source across all criteria, which was most pronounced for accuracy and perceived reliability. The divergence between lay and expert raters underscores the strong influence of evaluator background on communication-focused assessments.

These findings are broadly consistent with prior studies examining LLM performance in patient-oriented medical communication, while also highlighting important nuances. In a comparable study evaluating responses to rheumatology-related patient questions, both rheumatologists and patients assessed LLM-generated and clinician-generated answers. Although patient ratings did not differ significantly between the two approaches in terms of comprehensiveness or readability, rheumatologists rated LLM responses lower with respect to accuracy and completeness, and preference rates differed markedly between patients and experts[13]. In contrast, our study found higher ratings for LLM-generated responses, particularly among layperson raters, whereas differences within the expert subgroup were less pronounced. Notably, no statistically significant differences were found in clarity scores between LLM and physician responses.

Table 3a. Bonferroni-adjusted pairwise comparisons (p)-all raters			
Comparison	**A/PR (p)**	**C (p)**	**SA/PU (p)**
Claude Sonnet 4 vs. DeepSeek v3.2	1.000	1.000	1.000
Claude Sonnet 4 vs. GPT-4	0.349	1.000	0.752
Claude Sonnet 4 vs. Gemini 2.5	**<0.001**	0.365	**<0.001**
Claude Sonnet 4 vs. Professor	1.000	**0.010**	0.541
DeepSeek v3.2 vs. GPT-4	0.365	**0.025**	1.000
DeepSeek v3.2 vs. Gemini 2.5	**<0.001**	**0.021**	**0.021**
DeepSeek v3.2 vs. Professor	1.000	0.218	0.136
GPT-4 vs. Gemini 2.5	**0.009**	1.000	**0.011**
GPT-4 vs. Professor	**<0.001**	**<0.001**	**0.015**
Gemini 2.5 vs. Professor	**<0.001**	**<0.001**	**<0.001**
A/PR: Accuracy (experts)/perceived reliability (laypersons), C: Clarity, SA/PU: Scientific adequacy (experts)/practical usefulness (laypersons)			

Table 3b. Bonferroni-adjusted pairwise comparisons (p)-experts only			
Comparison	**A/PR (p)**	**C (p)**	**SA/PU (p)**
Claude Sonnet 4 vs. DeepSeek v3.2	0.789	1.000	0.130
Claude Sonnet 4 vs. GPT-4	1.000	1.000	1.000
Claude Sonnet 4 vs. Gemini 2.5	**0.010**	1.000	**<0.001**
Claude Sonnet 4 vs. Professor	**0.046**	1.000	**0.004**
DeepSeek v3.2 vs. GPT-4	1.000	1.000	1.000
DeepSeek v3.2 vs. Gemini 2.5	0.404	1.000	0.699
DeepSeek v3.2 vs. Professor	1.000	1.000	1.000
GPT-4 vs. Gemini 2.5	0.095	1.000	**0.026**
GPT-4 vs. Professor	0.132	1.000	0.284
Gemini 2.5 vs. Professor	1.000	1.000	1.000
A/PR: Accuracy (experts)/perceived reliability (laypersons), C: Clarity, SA/PU: Scientific adequacy (experts)/practical usefulness (laypersons)			

Table 3c. Bonferroni-adjusted pairwise comparisons (p)-laypersons only			
Comparison	**A/PR (p)**	**C (p)**	**SA/PU (p)**
Claude Sonnet 4 vs. DeepSeek v3.2	0.844	0.595	1.000
Claude Sonnet 4 vs. GPT-4	0.351	1.000	1.000
Claude Sonnet 4 vs. Gemini 2.5	**<0.001**	0.628	**0.036**
Claude Sonnet 4 vs. Professor	**<0.001**	**<0.001**	**<0.001**
DeepSeek v3.2 vs. GPT-4	**0.021**	**0.019**	0.814
DeepSeek v3.2 vs. Gemini 2.5	**<0.001**	**0.019**	**0.013**
DeepSeek v3.2 vs. Professor	**0.007**	**0.016**	**<0.001**
GPT-4 vs. Gemini 2.5	**0.040**	1.000	1.000
GPT-4 vs. Professor	**<0.001**	**<0.001**	**<0.001**
Gemini 2.5 vs. Professor	**<0.001**	**<0.001**	**<0.001**
Bold p-values indicate p<0.05, (p) computed on per-question means (n=37) with Bonferroni correction A/PR: Accuracy (experts)/perceived reliability (laypersons), C: Clarity, SA/PU: Scientific adequacy (experts)/practical usefulness (laypersons)			

Similar patterns have been reported in other clinical contexts. An emergency medicine study comparing LLM-generated patient education responses with those produced by orthopedic clinicians found more favorable ratings for AI-generated content across multiple domains, aligning with the communication-related advantages observed in our study[18]. Likewise, investigations focusing on patient questions related to lung cancer surgery reported high ratings for both accuracy and clarity of LLM responses, while emphasizing that such systems should be used as educational or supportive tools rather than as independent clinical decision-makers[15].

In contrast, studies assessing diagnostic and triage performance present a more cautious perspective. Evaluations comparing LLMs with physicians and laypersons have shown that while LLMs may outperform non-experts and provide information perceived as more useful than conventional search engine outputs, they continue to underperform physicians in diagnostic accuracy and complex clinical reasoning tasks[19]. Similarly, a recent systematic review reported that although LLMs can achieve accuracy comparable to specialists in selected imaging-related tasks, they remain limited in patient-centered communication and clinical reasoning, reinforcing the ongoing necessity of human clinical expertise[11]. Consistent with these observations, Liu et al.[20] demonstrated that LLMs performed below senior physicians in pulmonary diagnostic tasks, highlighting the distinction between theoretical knowledge synthesis and integrative, experience-based diagnostic reasoning.

Beyond comparisons between physicians and AI systems, our findings also point to differences among LLMs. Although Gemini 2.5 achieved higher mean ratings in several domains, these differences did not consistently reach statistical significance in assessments by expert evaluators. Among laypersons, however, Gemini 2.5 was rated more favorably than other models, suggesting that communicative fluency and interpretability may play a stronger role in public perception than domain-specific accuracy alone. This observation aligns with findings from intensive care medicine, where GPT-4-based models demonstrated high accuracy but were also associated with increased computational demands and concerns regarding confidently incorrect outputs, underscoring the importance of ongoing validation prior to broader clinical adoption[21].

Multiple evaluations of medical AI systems have further identified persistent challenges, including logical inconsistencies, contextual fragmentation, and limitations in deep diagnostic reasoning[22,23]. As Turner[24] has noted, trust remains an ethical concept that is difficult to ascribe to algorithmic systems lacking moral agency. In this context, the present findings support the view that LLMs should be regarded as adjunctive tools designed to augment- rather than replace-physicians. Their greatest potential lies in supporting medical communication and information delivery, while human clinicians remain essential for clinical judgment, accountability, empathy, and holistic decision-making at the core of medical practice.

Study Limitations

This study has several limitations. First, physician responses were obtained from a single pulmonologist, which may limit the generalizability of the findings to broader clinical practice. Second, the number of evaluators was limited, restricting the extent to which the results can be considered representative of larger expert and lay populations. Third, all primary outcome measures were based on subjective ratings of communication quality rather than on objective assessments of factual correctness or clinical validity. In addition, although prompts were predefined and AI-generated responses were archived at the time of evaluation, LLMs are continuously updated, which may affect reproducibility across different model versions. The study focused on perceived communication quality and the usefulness of information; therefore, the findings should not be interpreted as reflecting diagnostic accuracy, clinical reasoning performance, or decision-making capability. In addition, perceived accuracy was rated by humans without gold standard (e.g., guideline-based) adjudication of factual correctness; the human comparator was a single pulmonologist, limiting generalizability to physicians as a group; the question set did not undergo formal content validation; response length differed across sources and was positively associated with ratings, representing a potential format confounder; repeated-measures ANOVA, rather than mixed-effects modeling, was used for the primary comparisons. The linguistic and contextual characteristics of the questions and evaluators may also limit generalizability to other languages or health care settings.

Conclusion

This study suggests that current LLMs can generate pulmonology-related medical information that is perceived as clear and scientifically adequate by both expert and non-expert audiences. Although some models received higher

communication-related ratings, these results should be interpreted in light of the subjective nature of the evaluations and do not imply equivalence to clinical expertise. The advantage of LLMs was driven mainly by lay raters; among expert raters, the single pulmonologist was rated at least as highly as the strongest models for accuracy and scientific adequacy.

Importantly, the results underscore that LLMs are best viewed as supportive tools rather than as substitutes for physicians. Human clinicians remain essential for clinical reasoning, contextual judgment, empathy, and professional accountability. Accordingly, the integration of AI-assisted information delivery with physician oversight may enhance patient communication while preserving the central role of human expertise in medical practice.

Ethics

Ethics Committee Approval: This study did not involve patients, patient data, identifiable personal information, biological samples, or any clinical intervention. The raters evaluated anonymized written responses; therefore, ethics committee approval was not required according to institutional and national regulations.

Informed Consent: Not applicable, as the study did not involve patients or identifiable personal data.

Footnotes

Authorship Contributions

Surgical and Medical Practices: D.S.U., Ö.S.U., M.O.G., A.A., A.E.E., Concept: D.S.U., Ö.S.U., M.O.G., Design: D.S.U., Ö.S.U., Data Collection or Processing: D.S.U., Ö.S.U., B.C.K., A.A., Analysis or Interpretation: D.S.U., Ö.S.U., B.C.K., A.A., Literature Search: D.S.U., M.O.G., B.C.K., A.E.E., Writing: D.S.U., Ö.S.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.

Supplementary Tables: https://d2v96fxpocvxx.cloudfront.net/cf9d60d6-523c-458a-a2e6-78728d3ffbb0/content-images/e99fe721-803e-4f4f-9249-125ccbd841b5.pdf

References

1. Lie HC, Juvet LK, Street RL Jr, et al. Effects of physicians' information giving on patient outcomes: a systematic review. J Gen Intern Med. 2022;37:651-63.
2. Zakaria M, Mazumder S, Faisal HM, et al. Physician communication behaviors on patient satisfaction in primary care medical settings in Bangladesh. J Prim Care Community Health. 2024;15:21501319241277396.
3. Sharkiya SH. Quality communication can improve patient-centred health outcomes among older patients: a rapid review. BMC Health Serv Res. 2023;23:886.
4. Jia X, Pang Y, Liu LS. Online health information seeking behavior: a systematic review. Healthcare (Basel). 2021;9:1740.
5. Daraz L, Dogu C, Houde V, Bouseh S, Morshed KG. Assessing credibility: quality criteria for patients, caregivers, and the public in online health information-a qualitative study. J Patient Exp. 2024;11:23743735241259440.
6. Sonig A, Deeney C, Hurley ME, et al. What patients and caregivers want to know when consenting to the use of digital behavioral markers. NPP Digit Psychiatry Neurosci. 2024;2:19.
7. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. Radiographics. 2017;37:505-15.
8. Neri E, Coppola F, Miele V, Bibbolino C, Grassi R. Artificial intelligence: who is responsible for the diagnosis? Radiol Med. 2020;125:517-21.
9. Beaulieu-Jones BK, Greene CS. Reproducibility of computational workflows is automated using continuous analysis. Nat Biotechnol. 2017;35:342-6.
10. Hosny A, Parmar C, Coroller TP, et al. Deep learning for lung cancer prognostication: a retrospective multi-cohort radiomics study. PLoS Med. 2018;15:e1002711.
11. Shen J, Zhang CJP, Jiang B, et al. Artificial intelligence versus clinicians in disease diagnosis: systematic review. JMIR Med Inform. 2019;7:e10010.
12. Kusal S, Patil S, Choudrie J, Kotecha K, Mishra S, Abraham A. AI-based conversational agents: a scoping review from technologies to future directions. IEEE Access. 2022;10:92337-56.
13. Ye C, Zweck E, Ma Z, Smith J, Katz S. Doctor versus artificial intelligence: patient and physician evaluation of large language model responses to rheumatology patient questions in a cross-sectional study. Arthritis Rheumatol. 2024;76:479-84.
14. Charide R, Stallwood L, Munan M, et al. Knowledge mobilization activities to support decision-making by youth, parents, and adults using a systematic and living map of evidence and recommendations on COVID-19: protocol for three randomized controlled trials and qualitative user-experience studies. Trials. 2023;24:27.
15. Akçay O, Öztürk Ö, Acar T, Gürsoy S. Accuracy and reliability of ChatGPT in answering patient questions about lung cancer and its surgery: an expert panel evaluation by thoracic surgeons. J Cancer Educ. 2026;41:477-82.
16. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15:155-63.
17. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977;33:159-74.
18. Liu J, Segal K, Daher M, et al. Artificial intelligence versus orthopedic surgeons as an orthopedic consultant in the emergency department. Injury. 2025;56:112297.

19. Levine DM, Tuwani R, Kompa B, et al. The diagnostic and triage accuracy of the GPT-3 artificial intelligence model. medRxiv [Preprint]. 2023:2023.01.30.23285067.
20. Liu X, Liu H, Yang G, et al. A generalist medical language model for disease diagnosis assistance. Nat Med. 2025;31:932-42.
21. Workum JD, Volkers BWS, van de Sande D, et al. Comparative evaluation and performance of large language models on expert level critical care questions: a benchmark study. Crit Care. 2025;29:72.
22. Hager P, Jungmann F, Holland R, et al. Evaluation and mitigation of the limitations of large language models in clinical decision-making. Nat Med. 2024;30:2613-22.
23. Griot M, Hemptinne C, Vanderdonckt J, Yuksel D. Large language models lack essential metacognition for reliable medical reasoning. Nat Commun. 2025;16:642.
24. Turner JH. Triangle of trust in cancer care? The physician, the patient, and artificial intelligence chatbot. Cancer Biother Radiopharm. 2023;38:581-4.

DOI: 10.4274/anatoljmed.2026.99815
Anatol J Gen Med Res 2026;36(2):212-21

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Hemodynamic Effects of an Augmented Low-dose Dexamethasone Protocol in Preterm Infants with Bronchopulmonary Dysplasia

Bronkopulmoner Displazili Preterm Bebeklerde Artırılmış Düşük Doz Deksametazon Protokolünün Hemodinamik Etkileri

Oğuzhan Ay[1], Sezgin Güneş[2], Nurcan Aydoğan[3], Ayşe Şimşek[1,4]

[1]İzmir Democracy University Buca Seyfi Demirsoy Training and Research Hospital, Department of Pediatrics, Division of Pediatric Cardiology, İzmir, Türkiye
[2]University of Health Sciences Türkiye, İzmir City Hospital, Department of Neonatology, İzmir, Türkiye
[3]İzmir Democracy University Buca Seyfi Demirsoy Training and Research Hospital, Department of Pediatrics, İzmir, Türkiye
[4]İzmir Democracy University Faculty of Medicine, Department of Pediatrics, İzmir, Türkiye

Cite as: Ay O, Güneş S, Aydoğan N, Şimşek A. Hemodynamic effects of an augmented low-dose dexamethasone protocol in preterm infants with bronchopulmonary dysplasia. Anatol J Gen Med Res. 2026;36(2):212-21

Abstract

Objective: This study aimed to investigate the cardiovascular impact of an augmented dexamethasone protocol in preterm infants with bronchopulmonary dysplasia (BPD), specifically evaluating changes in hemodynamic parameters before and after treatment.

Methods: A retrospective analysis was conducted on 20 preterm infants (median gestational age: 28 weeks; median birth weight: 1245 g) who received an augmented steroid regimen. Hemodynamic data, including heart rate and arterial blood pressure, were collected from neonatal intensive care unit monitoring and were analyzed using the Wilcoxon signed-rank test. Demographic characteristics were compared with established literature to ensure that the cohort was representative of the high-risk BPD population.

Results: Statistical analysis revealed a significant elevation in systolic blood pressure following steroid administration ($p=0.006$, $Z=1.932$, $r=0.43$). In contrast, no statistically significant differences were observed in heart rate ($p=0.131$), diastolic blood pressure ($p=0.872$), and mean arterial pressure ($p=0.228$). While existing literature often reports a higher risk of BPD and mortality in male infants, this study found no significant gender-based differences in clinical characteristics or in treatment response, suggesting a homogeneous response to the standardized protocol across the cohort.

Conclusion: The findings demonstrate that an augmented dexamethasone protocol is associated with a significant increase in systolic blood pressure, underscoring the dose-dependent nature of corticosteroid-induced cardiovascular adverse effects. Given the increased risk of systemic hypertension, it is critically important to prioritize dose optimization and implement rigorous hemodynamic monitoring during therapy. Prospective trials are essential to refine personalized treatment strategies and to ensure the long-term cardiovascular safety of preterm infants receiving steroid treatment for BPD.

Keywords: Bronchopulmonary dysplasia, preterm infants, dexamethasone, systemic hypertension

Address for Correspondence/Yazışma Adresi: Oğuzhan Ay, MD, İzmir Democracy University Buca Seyfi Demirsoy Training and Research Hospital, Department of Pediatrics, Division of Pediatric Cardiology, İzmir, Türkiye
E-mail: oguzhanay1@gmail.com
ORCID ID: orcid.org/0000-0002-8356-4113

Received/Geliş tarihi: 24.06.2026
Accepted/Kabul tarihi: 25.07.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

Amaç: Bu çalışma, bronkopulmoner displazisi (BPD) olan prematüre bebeklerde artırılmış düşük doz deksametazon protokolünün kardiyovasküler etkilerini araştırmayı hedeflemiştir; özellikle tedavi öncesi ve sonrası hemodinamik parametrelerdeki değişiklikleri değerlendirerek.

Yöntem: Artırılmış steroid rejimi alan 20 prematüre bebekte geriye dönük bir analiz gerçekleştirildi (ortalama gebelik süresi: 28 hafta; ortalama doğum ağırlığı: 1245 g). Kalp atış hızı ve arteriyel kan basıncı gibi hemodinamik veriler, yenidoğan yoğun bakım ünitesi izlemiyle toplanarak Wilcoxon imzalı rank testi ile analiz edildi. Kohortun yüksek riskli BPD popülasyonunu temsil ettiğinden emin olmak amacıyla demografik özellikler, mevcut literatürle karşılaştırılmıştır.

Bulgular: İstatistiksel analiz, steroid uygulamasından sonra sistolik kan basıncında anlamlı bir yükselme gösterdi (p=0,006, Z=1,932, r=0,43). Buna karşılık, kalp atış hızında (p=0,131), diyastolik kan basıncında (p=0,872) ve ortalama arteryel basınçta (p=0,228) istatistiksel olarak anlamlı farklar gözlenmedi. Mevcut literatür erkek bebeklerde BPD ve mortalite riskinin daha yüksek olduğunu bildirirken, bu çalışmada klinik özelliklerde veya tedavi yanıtında anlamlı cinsiyet farkları bulunmadı; standardize protokole homojen kohort yanıtı öneriliyor.

Sonuç: Bulgular, artırılmış deksametazon protokolünün sistolik kan basıncında dikkate değer ve anlamlı bir artışla ilişkili olduğunu göstermektedir; bu, kortikosteroid kaynaklı kardiyovasküler yan etkilerin doza bağlı doğasını vurgulamaktadır. Sistemik hipertansiyon riskinin artması nedeniyle, doz optimizasyonuna öncelik verilmesi ve tedavi sırasında titiz hemodinamik izleme yapılması kritik öneme sahiptir. Gelecekteki prospektif çalışmalar, kişiselleştirilmiş tedavi stratejilerini rafine etmek ve BPD için steroid tedavisi gören prematüre bebeklerin uzun vadeli kardiyovasküler güvenliğini sağlamak amacıyla zorunludur.

Anahtar Kelimeler: Bronkopulmoner displazi, prematüre bebekler, deksametazon, sistemik hipertansiyon

Introduction

Prematurity, especially when associated with serious respiratory disorders such as bronchopulmonary dysplasia (BPD), is a significant cause of morbidity and mortality in neonatal intensive care units[1]. In these cases, steroid therapy for developing BPD is an accepted primary treatment aimed at improving respiratory status and facilitating extubation[2,3]. However, ongoing research on corticosteroid dosages and their hemodynamic side effects is critically important for determining optimal treatment strategies[2,4]. In this context, the potential effects of steroid treatment administered at doses higher than those used in the low-dose steroid protocol on the cardiac function in premature infants, particularly regarding long-term outcomes, require in-depth examination. Hemodynamic dysfunction is a common complication, and steroid interactions during this sensitive period may lead to revisions to treatment protocols[1,5]. In extremely preterm newborns, significant respiratory distress due to impaired pulmonary compliance despite surfactant treatment, together with interventions such as mechanical ventilation and hyperoxic gas mixtures, increases susceptibility to chronic pulmonary disorders[6,7]. Although the role of corticosteroids in preventing and treating these chronic disorders is undisputed, the hemodynamic risks associated with optimal dosage and timing of administration have not been fully elucidated[6,8]. Although synthetic glucocorticoids, such as dexamethasone, have been shown to be effective in reducing the risk of BPD in preterm infants, high-dose use is associated with serious cardiac side effects, such as hypertension and hypertrophic cardiomyopathy[2,6]. Therefore, the effects of steroid doses used to treat BPD in premature infants on the cardiac system should be evaluated in detail with respect to both short- and long-term outcomes. In the literature, systemic steroid treatment regimens for BPD are classified as low-dose (cumulative dose under 2 mg/kg), medium-dose (2-4 mg/kg), and high-dose (4-8 mg/kg)[2]. Studies related to dexamethasone: a randomized trial (DART) and augmented low-dose systemic steroid treatment protocols are available in the literature[1,8,9]. This evaluation will contribute to treatment guidelines by retrospectively examining the hemodynamic side effects of the augmented low-dose protocol in preterm infants, which falls within the low-dose systemic steroid treatment regimen, based on echocardiograms, electrocardiograms, and neonatal intensive care follow-up[8]. This study specifically aims to identify adverse hemodynamic effects[10,11]. Evidence indicates that antenatal glucocorticoids accelerate the maturation of the fetal respiratory and cardiovascular systems and are effective in preventing and treating postnatal respiratory distress syndrome[12]. However, the short- and long-term effects of these treatments, especially antenatal maternal corticosteroid use in cases of high-risk preterm birth, on fetal circulation and cardiac function are not fully understood[13]. The benefits of postnatal steroid treatment, such as accelerating weaning from mechanical ventilation and improving respiratory outcomes in preterm infants with BPD, have been demonstrated clinically[6]. Nevertheless, these

benefits must be carefully balanced against possible short- and long-term side effects such as systemic hypertension, hyperglycemia, growth retardation, gastrointestinal bleeding, intestinal perforation, and hypertrophic cardiomyopathy[8,14]. Therefore, the optimal dose and duration of postnatal steroid treatment should be meticulously determined, considering both pulmonary benefits and cardiac risks[8,15].

Specifically, the effects of steroids administered at doses higher than those used in the DART protocol on myocardial growth, functional changes, and vascular remodeling in preterm infants need to be comprehensively examined[6]. In this context, the maturational effects of antenatal glucocorticoid treatment on fetal heart structure and energy metabolism should be compared with the hemodynamic outcomes of postnatal steroid treatment, and the impacts of different dosage regimens on the cardiovascular system should be detailed[16]. Additionally, evaluating the potential effects of such dose differences on hemodynamic parameters such as myocardial hypertrophy, changes in hemodynamic output, and vascular resistance will provide valuable insights for clinical practice. Moreover, the effects of corticosteroid use in preterm infants on long-term cardiovascular outcomes represent a critical research area for assessing the risks of cardiometabolic diseases in adulthood[8,17]. Given cortisol's biological roles and the potential of exogenous hydrocortisone to prevent complications of adrenal insufficiency in preterm infants, the hemodynamic effects of different steroid types should be compared[18]. From this perspective, our study will enhance understanding of the short- and long-term effects of varying steroid dosages on preterm infants' cardiac health, thereby contributing to the optimization of clinical treatment approaches[6].

In this study, the hemodynamic effects of augmented low-dose systemic steroid treatment, compared with the DART protocol, were retrospectively examined in premature infants who developed BPD. The hemodynamic effects of these groups were evaluated using echocardiography, electrocardiography, and follow-up in the neonatal intensive care unit. This study aimed to characterize pre- and post-treatment hemodynamic adverse effects of dexamethasone in preterm infants diagnosed with BPD who had no echocardiographic evidence of hypertrophic cardiomyopathy or ventricular wall thickening. This approach will contribute to treatment optimization by providing clinical evidence for discussions of dosage in the literature.

Materials and Methods

Data Collection and Study Design

20 premature infants diagnosed with BPD and treated with augmented low-dose systemic steroid therapy according to the DART protocol in the Neonatal Intensive Care Unit of İzmir Democracy University Buca Seyfi Demirsoy Training and Research Hospital, within the last five years, were included in the retrospective cohort study. Data were collected by reviewing epicrisis reports, patient files, and electronic records. The collected data encompassed prenatal, perinatal, and postnatal histories; maternal age; gestational age; birth weight; Apgar scores; duration of hospital stay; steroid initiation day and duration; durations of respiratory support; discharge weight; oxygen requirement at 36 weeks; pre- and post-steroid heart rates and systolic and diastolic blood pressures; tricuspid regurgitation and right ventricular systolic pressure; physical examination findings; medical and family history; vital signs; and electrocardiogram and echocardiography reports. Post-steroid comprised measurements obtained within 24 hours of completing the 10-day steroid treatment.

A neonatal intensive care specialist diagnosed BPD according to the Turkish Neonatology Guideline and arranged steroid and other treatments provided in the intensive care unit in accordance with the current guideline[19,20].

Study Steroid Treatment Protocol

The DART protocol is a low-dose dexamethasone regimen developed to facilitate extubation and prevent BPD in ventilator- and oxygen-dependent preterm infants who have developed BPD or are at risk of it[1,2]. This protocol involves intravenous administration over 10 days, consisting of tapering doses targeting a total cumulative dose of 0.89 mg/kg of dexamethasone[2,21]. Our study included 20 preterm infants diagnosed with BPD who received an augmented low-dose steroid treatment that differed from the DART protocol. The median cumulative steroid dose administered to these patients was 1.35 mg/kg. We verified individualized cumulative doses for all participants to ensure adherence to the protocol; we excluded infants receiving non-standard dosages. The institutional low-dose dexamethasone regimen involved intravenous administration of 0.2 mg/kg/day (days 1-3), 0.15 mg/kg/day (days 4-6), and 0.1 mg/kg/day (days 7-9)[1,6].

Exclusion Criteria

Patients not included in the study were excluded according to the following criteria:

• Premature infants treated with a median cumulative steroid dose other than 1.35 mg/kg.

• Patients with missing prenatal, perinatal, postnatal history, physical examination, vital signs, echocardiography data in their files.

• Patients in whom congenital heart disease was detected on echocardiography.

• Patients diagnosed with conditions other than BPD according to the Turkish Neonatology Society guideline.

• Patients with chromosomal anomalies, major congenital malformations, or congenital surfactant deficiency.

• Patients receiving alternative systemic steroid treatments (e.g., hydrocortisone), as well as neonates receiving inotropes with known hemodynamic effects. Patients with high gestational age/birth weight.

• Patients who developed acute kidney injury-defined as a serum creatinine rise ≥0.3 mg/dL within 48 hours or oliguria (<1 ml/kg/h) for ≥24 hours during neonatal intensive care unit follow-up-were excluded from the study[22].

Statistical Analysis

Statistical analyses were performed using the open-source software JASP (version 0.95.4)[23]. Continuous variables were expressed as mean ± standard deviation if normally distributed or as median (minimum-maximum) if not. Categorical variables were presented as numbers (percentages). Normality of continuous data was assessed using the Kolmogorov-Smirnov test. Bivariate associations between continuous variables were evaluated using Pearson's correlation coefficient for normally distributed data and Spearman's rank correlation coefficient for non-normally distributed data. Differences between categorical variables were investigated using the chi-square test (or Fisher's exact test if expected cell counts were <5). Comparisons of quantitative variables between independent groups were performed using Student's independent t-test or for non-normally distributed parameters. Paired pre- and post-steroid comparisons were analyzed using paired t-tests or Wilcoxon signed-rank tests, as appropriate. A p-value <0.05 was considered statistically significant for all tests.

Ethics Committee Approval

Ethical approval for this study was obtained from the İzmir Democracy University Buca Seyfi Demirsoy Training and Research Hospital Non-Interventional Research Ethics Committee (approval no: 2025/530, date: 26.11.2025). The study was conducted in accordance with the ethical principles of the Declaration of Helsinki, and the committee approved a waiver of informed consent for the retrospective analysis of anonymized patient data.

Results

Of the patients included in the study, 4 were born via normal vaginal delivery, while the other 16 were born via cesarean section. The prenatal diagnoses were as follows: gestational diabetes in 2 patients, premature rupture of membranes in 2 patients, maternal hypertension in 2 patients, and chorioamnionitis in 1 patient. Surfactant was administered to 15 patients.

According to the echocardiography reports, the diagnoses documented in the patients' final echocardiograms prior to discharge were as follows: in 1 patient, surgical ligation had been performed for a hemodynamically significant patent ductus arteriosus (PDA) 2 weeks prior to dexamethasone initiation; a hemodynamically insignificant PDA was present in 2 patients; a medium-sized (5-6 mm) secundum atrial septal defect was present in 2 patients; and pulmonary hypertension was not detected in any of the patients who were evaluated with pre- and- post-steroid echocardiography. Echocardiographic evaluation revealed no evidence of hypertrophic cardiomyopathy, interventricular septal thickening, or ventricular wall thickening in any of the patients.

Analysis of the demographic and clinical characteristics of the 20 preterm infants included in our study revealed a median maternal age of 27 years (range: 19-39). The median gestational age was 28 weeks (range: 21-32 weeks), and the median birth weight was 1245 grams (range: 620-1925 grams). The median duration of hospital stay was determined to be 52 days (range: 16-120). The median time to initiation of steroid therapy was 29 days (range: 6-59 days). The median discharge weight was 2800 grams (range: 1805-3130) (Table 1).

Throughout the follow-up period, the mechanical ventilation parameters of enrolled patients were systematically recorded and included in the analysis, whereas incubator and nasal-cannula oxygen support were not included in the quantitative

Table 1. Demographic and clinical characteristics of the preterm infants included in the study			
Variables	**Median**	**Minimum**	**Maximum**
Maternal age (years)	27	19	39
Gestational age (weeks)	28	21	32
Birth weight (g)	1245	620	1925
Length of hospital stay (days)	52	16	120
Postnatal age at steroid initiation (days)	29	6	59
1-minute Apgar score	5	1	8
5-minute Apgar score	7	5	8
Duration of invasive mechanical ventilation (days)	9	3	91
Duration of non-invasive mechanical ventilation (days)	15	0	59
Total duration of invasive and non-invasive respiratory support (days)	45	12	91
Weight at discharge (g)	2800	1805	3130
Oxygen saturation at 36 weeks (%)	97	90	100
Pre-steroid heart rate (bpm)	154	133	168
Post-steroid heart rate (bpm)	142	128	180
Pre-steroid systolic pressure (mmHg)	79	71	96
Pre-streoid diastolic pressure (mmHg)	43	22	52
Pre-steroid mean arterial pressure (mmHg)	56	39	65
Post-steroid systolic pressure (mmHg)	87	70	95
Post steroid diastolic pressure (mmHg)	38	35	55
Post-steroid mean arterial pressure (mmHg)	55	47	67

evaluation. Regarding the duration of respiratory support, the median duration of invasive mechanical ventilation was 9 days (range: 3-91), and that of non-invasive mechanical ventilation was 15 days (range: 0-59). The median total duration of respiratory support was 45 days (range: 12-91 days). At the 36th week, the median oxygenation level was 97% (range: 90-100%) (Table 1).

Upon evaluation of hemodynamic parameters, the median pre-steroid heart rate was 154 (range: 133-168) beats per minute, whereas the median post-steroid heart rate was 142 (range: 128-180) beats per minute. The median pre-steroid systolic and diastolic blood pressures were 79 (range: 71-96) mmHg and 43 (range: 22-52) mmHg, respectively. Following steroid treatment, the medians were 87 (range: 70-95) mmHg for systolic and 38 (range: 35-55) mmHg for diastolic pressure. Pre-steroid mean arterial pressure 56 (range: 39-65) mmHg and post-steroid mean arterial pressure 55 (range: 47-67) mmHg. Median Apgar scores were 5 (range, 1-8) at 1 minute and 7 (range, 5-8) at 5 minutes (Table 1).

Demographic and Clinical Characteristics by Gender

Demographic and clinical characteristics by gender are presented in detail in Table 2. Upon analysis of the demographic data, the median maternal age was 27 years among mothers of female neonates and 28 years among mothers of male neonates (p=0.756). The median gestational age was 28 weeks for female infants and 26 weeks for male infants (p=0.150), while the median birth weight was 1395 g for female infants and 1190 g for male infants (p=0.305). Median Apgar scores at the 1st minute were 5 for females and 4 for males (p=1.0); at the 5th minute, they were 7 for females and 6 for males (p=0.608; Table 2).

The median length of hospital stay was 62 days for female neonates and 46 days for male neonates (p=0.463). Additionally, the median postnatal age at the initiation of steroid therapy was 35 days in females and 25 days in males (p=0.096).

In terms of respiratory support durations, the median duration of invasive mechanical ventilation was 8 days in females and

Table 2. Analysis of demographic variables and clinical outcomes between male and female infants

Variables	Female neonates (n=8) median (Q1-Q3)	Male neonates (n=12) median (Q1-Q3)	p-value
Maternal age (years)	27 (21-31)	28 (22-30)	0.756
Gestational age (weeks)	28 (27-30)	26 (24-28)	0.150
Birth weight (gr)	1395 (1163-1563)	1190 (767-1360)	0.305
1-minute Apgar score	5 (4-6)	4 (3-6)	1.0
5-minute Apgar score	7 (6-8)	6 (6-7)	0.608
Length of hospital stay (days)	62 (46-72)	46 (36-71)	0.463
Postnatal age at steroid initiation (days)	35 (29-44)	25 (15-30)	0.096
Duration of invasive mechanical ventilation (days)	8 (5-11)	14 (7-25)	0.332
Duration of non-invasive mechanical ventilation (days)	12 (5-28)	17 (0-41)	0.814
Total duration of respiratory support (days)	30 (22-43)	63 (24-69)	0.111
Weight at discharge (g)	2860 (2680-2953)	2770 (2504-2903)	0.694
Oxygen saturation at 36 weeks (%)	98 (97-99)	97 (96-98)	0.546
Pre-steroid heart rate (bpm)	158 (151-165)	153 (138-155)	0.069
Post-steroid heart rate (bpm)	154 (140-161)	139 (133-149)	0.960
Pre-steroid systolic blood pressure (mmHg)	79 (72-85)	79 (73-85)	1.0
Pre-steroid diastolic blood pressure (mmHg)	43 (41-47)	42 (41-48)	0.938
Pre-steroid mean arterial pressure (mmHg)	57 (54-58)	55 (51-58)	0.508
Post-steroid systolic blood pressure (mmHg)	85 (71-89)	88 (83-92)	0.437
Post-steroid diastolic blood pressure (mmHg)	38 (37-55)	38 (36-46)	0.504
Post-steroid mean arterial pressure (mmHg)	55 (48-64)	55 (53-57)	0.908

14 days in males (p=0.332); the median duration of non-invasive mechanical ventilation was 12 days in females and 17 days in males (p=0.814); and the median total duration of respiratory support was 30 days in females and 63 days in males (p=0.111). The median weight at discharge was 2860 g in females and 2770 g in males (p=0.694). Oxygen saturation at 36 weeks was 98% in females and 97% in males (p=0.546).

Upon examination of hemodynamic parameters, the median pre-steroid heart rates were 158 bpm in female neonates and 153 bpm in male neonates (p=0.069); the median post-steroid heart rates were 154 bpm in female neonates and 139 bpm in male neonates (p=0.960). The median pre-steroid systolic blood pressure was 79 mmHg in both females and males (p=1.0); the median pre-steroid diastolic blood pressure was 43 mmHg and 42 mmHg, respectively (p=0.938). Post-steroid systolic blood pressures were 85 mmHg in females and 88 mmHg in males (p=0.437), and diastolic blood pressures were 38 mmHg in both groups (p=0.504). Pre-steroid mean arterial pressure was 57 and 55 mmHg (p=0.508); post-steroid mean arterial pressure was 55 and 55 mmHg (p=0.908). No statistically significant differences were observed between genders across all parameters analyzed (Table 2).

Hemodynamic Findings Before and After Steroid Treatment

The impact of steroid therapy on hemodynamic parameters was evaluated using the Wilcoxon signed-rank test, a paired-samples test. No statistically significant difference was observed between pre- and post-steroid heart rates (p=0.131, Z=-1.529, r=0.34). In contrast, a significant change was noted when comparing pre- and post-steroid systolic blood pressure (p=0.006, Z=1.932, r=0.43), indicating an increase in systolic blood pressure following steroid administration. However, no statistically significant difference was detected between pre- and post-steroid diastolic blood pressure (p=0.872, Z=-1.181, r=0.04). Furthermore, mean arterial pressure values were analyzed to determine the differences between pre- and post-administration measurements (p=0.228, Z=0.765, r=0.20) (Table 3, Figures 1-4).

Table 3. Comparative analysis of hemodynamic parameters before and after steroid treatment

Pre-post steroid data	W	z	r	p-value
Pre-post steroid heart rate	133.00	-1.529	0.34	0.131
Pre-post steroid systolic pressure	47.00	1.932	0.43	0.006
Pre-post streoid diastolic pressure	99.50	-0.181	0.04	0.872
Pre-post steroid mean arterial pressure	76.00	-0.765	0.20	0.228

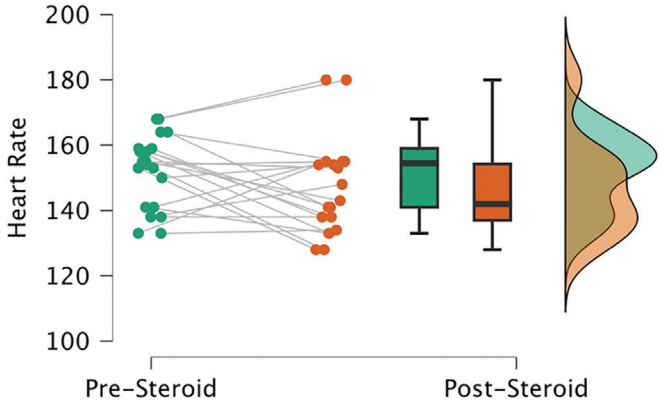


Figure 1. Distribution of heart rate before and after augmented dexamethasone treatment

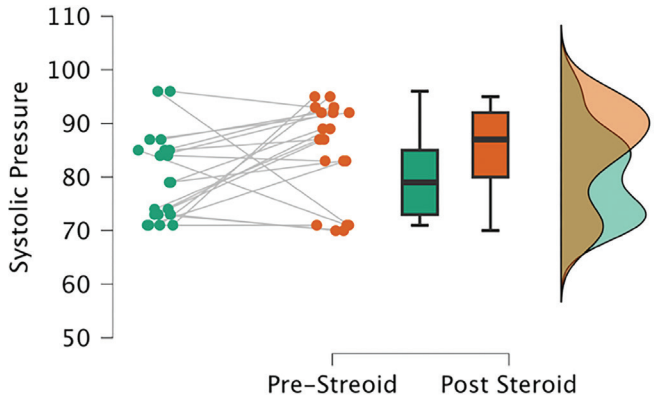


Figure 2. Distribution of systolic pressure before and after augmented dexamethasone treatment

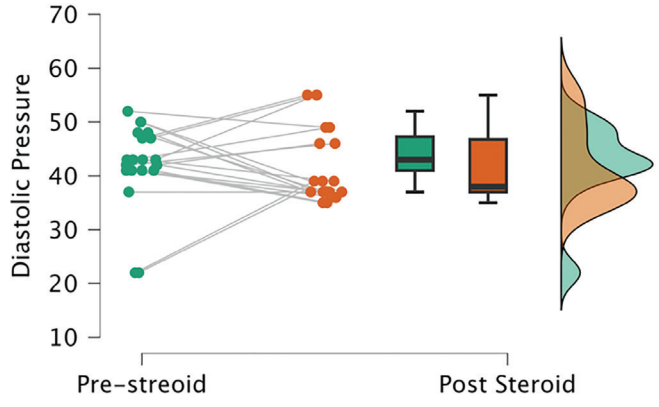


Figure 3. Distribution of diastolic pressure before and after augmented dexamethasone treatment

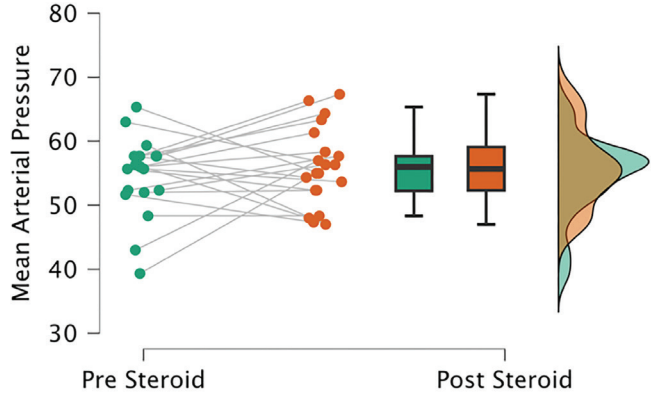


Figure 4. Distribution of mean arterial pressure before and after augmented dexamethasone treatment

Discussion

Demographic data in our study are consistent with those reported in the existing literature for preterm infants who develop BPD[2,11]. Within the general cohort, the median maternal age was 27 years. The median gestational age was 27 weeks for female infants and 28 weeks for male infants (overall median: 28 weeks). The median birth weight was 1395 g for females and 1190 g for males (overall median: 1245 g), while the median length of hospital stay was 62 days for females and 46 days for males (overall median: 52 days). These values parallel the demographic profiles typically reported in populations at risk for BPD; for example, gestational age <28 weeks and birth weight <1500 g are standard cohort criteria in systemic steroid studies[2,24].

As the differences between genders were not statistically significant, the cohort's homogeneity is consistent with that of similar retrospective and randomized controlled trial cohorts in the literature[25]. However, findings in the literature suggest that male preterm infants face higher risks of BPD and pulmonary hypertension and have higher mortality than female preterm infants[26,27]. Conversely, some studies indicate that female infants tend to have lower birth weights and are weaned from supplemental oxygen later than male infants[28]. In this context, the absence of significant differences between genders in BPD development or steroid treatment response in our study may reflect the unique clinical characteristics of this population (e.g., similar distributions of gestational age and birth weight) or the potential effects of a standardized treatment approach[29].

Furthermore, the literature reports that male preterm infants require corticosteroid therapy more frequently (23.6% in males vs. 20.1% in females) and receive more repeat courses (mean 1.67 in males vs. 1.59 in females)[29]. This discrepancy may be associated with hormonal differences or with the more severe disease progression often observed in males; therefore, the role of sex in BPD pathogenesis warrants further investigation to facilitate personalized treatments[29]. In this regard, a deeper understanding of the role of sex in BPD pathogenesis and treatment response is critical for developing personalized therapeutic strategies. Further research into the sex-specific effects of steroid dosages and timing could contribute to treatment optimization by elucidating the underlying biological mechanisms. Additionally, the significant change in systolic blood pressure observed in our study warrants detailed investigation of the cardiovascular regulatory mechanisms associated with steroid therapy in preterm infants.

Our study investigated the hemodynamic effects of corticosteroid therapy in preterm infants with BPD, specifically evaluating dose variations by pre- and post-administration measurements. The findings provide significant insights into the safety and efficacy of corticosteroid use in this population, particularly when compared with widely accepted low-dose strategies[30]. The slightly higher steroid dose used in our study resulted in a significant increase in systolic blood pressure[7,24]. This observation aligns with meta-analyses reporting an elevated risk of systemic hypertension in regimens in which the dexamethasone dose exceeds 0.25 mg/kg and supports the conclusion that the cardiovascular adverse effects of corticosteroids are dose-dependent[7,11,25]. Consequently, given that the risk of hypertension increases in regimens where the dexamethasone dose exceeds 0.25 mg/kg/day, it is critical to prioritize dose optimization, evaluate alternatives such as hydrocortisone that may carry lower cardiovascular risks, and elucidate long-term hemodynamic effects through prospective randomized controlled trials[2,7,11,24]. Advancing our understanding in these areas is essential for improving BPD outcomes and refining current therapeutic strategies[9,31].

In preterm infants with BPD, systemic steroid use does not reduce mortality; however, dexamethasone decreases the risk of BPD but may increase the risk of hypertension and hyperglycemia[2,11]. Such effects on hemodynamic parameters are particularly significant in studies comparing different dosage regimens, where high-dose protocols have been reported to shorten the duration of mechanical ventilation while potentially increasing adverse effects, such as hypertension and hyperglycemia[24]. In this context, the significant change in systolic blood pressure caused by our augmented low-dose steroid protocol highlights the delicate balance between dose optimization and minimizing hemodynamic side effects in BPD management. Indeed, studies evaluating changes in blood pressure and heart rate relative to corticosteroid dosages have demonstrated the direct impact of these pharmacological agents on the cardiovascular system[2,7,24]. This situation emphasizes that the benefit-risk balance of steroid use must be carefully evaluated, particularly in the management of BPD in preterm infants. Although the use of steroids in the treatment of BPD offers undeniable therapeutic benefits, individualizing dosage adjustments is critical to minimize potential adverse effects on hemodynamic function[32].

Study Limitations

This study is subject to several limitations. The small sample size and retrospective design primarily constrain the generalizability of the findings, owing to inherent challenges in conducting prospective studies among preterm infants. Additionally, the lack of long-term cardiovascular outcome assessments precludes definitive conclusions regarding the protocol's overall safety profile. Accordingly, prospective randomized controlled trials with larger cohorts and longer follow-up are essential to comprehensively evaluate the effects of augmented steroid dosages on clinical outcomes.

Conclusion

This study retrospectively evaluated the adverse hemodynamic effects of an augmented low-dose dexamethasone protocol in preterm infants who developed BPD, using echocardiography, electrocardiography, and neonatal intensive care unit follow-up data. Our findings, consistent with existing literature, demonstrate that increases in steroid dosages can lead to significant effects on blood pressure and may have long-term consequences for hemodynamic function. This retrospective analysis underscores the need for a more comprehensive investigation into the relationship between steroid dosage and hemodynamic adverse effects and further supports the evidence that dexamethasone elevates the risk of systemic hypertension. On the other hand, the observation that hydrocortisone does not consistently reduce BPD risk even at high doses, nor does it show a marked increase in cardiovascular adverse effects, suggests that the hemodynamic profiles of different corticosteroids may diverge.

Ethics

Ethics Committee Approval: Ethical approval for this study was obtained from the İzmir Democracy University Buca Seyfi Demirsoy Training and Research Hospital Non-Interventional Research Ethics Committee (approval no: 2025/530, date: 26.11.2025).

Informed Consent: The committee approved a waiver of informed consent for the retrospective analysis of anonymized patient data.

Footnotes

Authorship Contributions

Surgical and Medical Practices: O.Y., S.G., Concept: O.Y., S.G., A.Ş., Design: S.G., Data Collection or Processing: N.A., Analysis or Interpretation: O.Y., Literature Search: O.Y., S.G., Writing: O.Y.

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

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

References

1. Al-Taweel HM, Abdelhady ISI, Irfan N, et al. Comparing low-dose (DART) and enhanced low-dose dexamethasone regimens in preterm infants with bronchopulmonary dysplasia. Front Pediatr. 2023;11:1261316.
2. Raghuveer TS, Zackula RE, Lakhotia R, Binder SA. Systemic steroids and bronchopulmonary dysplasia: a systematic review and meta-analysis. J Perinatol. 2026;46:144-54.
3. Rosano K, Bery S, Marasch J, Farrell R, Mastrandrea LD, Ryan RM. Predicting iatrogenic adrenal insufficiency in neonates exposed to prolonged steroid courses: do cortisol levels help? J Perinatol. 2024;44:1658-62.
4. Sehgal A, Steenhorst JJ, Mclennan DI, Merkus D, Ivy D, McNamara PJ. The left heart, systemic circulation, and bronchopulmonary dysplasia: relevance to pathophysiology and therapeutics. J Pediatr. 2020;225:13-22.e2.
5. Huysman MW, Hokken-Koelega AC, De Ridder MA, Sauer PJ. Adrenal function in sick very preterm infants. Pediatr Res. 2000;48:629-33.
6. Vrselja A, Pillow JJ, Black MJ. Effect of preterm birth on cardiac and cardiomyocyte growth and the consequences of antenatal and postnatal glucocorticoid treatment. J Clin Med. 2021;10:3896.
7. Boscarino G, Cardilli V, Conti MG, et al. Outcomes of postnatal systemic corticosteroids administration in ventilated preterm newborns: a systematic review of randomized controlled trials. Front Pediatr. 2024;12:1344337.
8. Olaloko O, Mohammed R, Ojha U. Evaluating the use of corticosteroids in preventing and treating bronchopulmonary dysplasia in preterm neonates. Int J Gen Med. 2018;11:265-74.
9. Gilfillan MA, Kiladejo A, Bhandari V. Current and emerging therapies for prevention and treatment of bronchopulmonary dysplasia in preterm infants. Paediatr Drugs. 2025;27:539-62.
10. Esterman E, Goyen TA, Jani P, et al. Systemic postnatal corticosteroid use for the prevention of bronchopulmonary dysplasia and its relationship to early neurodevelopment in extremely preterm infants. World J Pediatr. 2023;19:586-94.
11. Szabó H, Baraldi E, Colin AA. Corticosteroids in the prevention and treatment of infants with bronchopulmonary dysplasia: part I. systemic corticosteroids. Pediatr Pulmonol. 2022;57:600-8.
12. Karlsson R, Kallio J, Toppari J, Scheinin M, Kero P. Antenatal and early postnatal dexamethasone treatment decreases cortisol secretion in preterm infants. Horm Res. 2000;53:170-6.
13. Henry A, Shand A, Welsh A. The short term fetal cardiovascular effects of corticosteroids used in obstetrics. Australas J Ultrasound Med. 2013;16:135-41.
14. Lardón M, Uberos J, Narbona E. Does corticosteroid treatment during the pre and postnatal periods affect the neurodevelopmental outcome of premature newborns? Biomedica. 2017;37:104-11.
15. Zhao C, He L, Li L, et al. Prenatal glucocorticoids exposure and adverse cardiovascular effects in offspring. Front Endocrinol (Lausanne). 2024;15:1430334.
16. Sakurai K, Takeba Y, Osada Y, et al. Antenatal glucocorticoid administration promotes cardiac structure and energy metabolism maturation in preterm fetuses. Int J Mol Sci. 2022;23:10186.
17. Ninan K, Liyanage SK, Murphy KE, Asztalos EV, McDonald SD. Evaluation of long-term outcomes associated with preterm exposure to antenatal corticosteroids: a systematic review and meta-analysis. JAMA Pediatr. 2022;176:e220483.
18. Jensen EA, Rysavy MA, Kusuda S; Tiny Baby Collaborative Steering Committee. Use of hydrocortisone in extremely preterm infants: emphasis on those born least mature. J Perinatol. 2025.
19. Northway WH Jr, Rosan RC, Porter DY. Pulmonary disease following respirator therapy of hyaline-membrane disease. Bronchopulmonary dysplasia. N Engl J Med. 1967;276:357-68.

20. Arsan S, Korkmaz A, Oğuz S. Turkish Neonatal Society guideline on prevention and management of bronchopulmonary dysplasia. Turk Pediatri Ars. 2018;53:S138-50.
21. Doyle LW, Davis PG, Morley CJ, McPhee A, Carlin JB; DART Study Investigators. Low-dose dexamethasone facilitates extubation among chronically ventilator-dependent infants: a multicenter, international, randomized, controlled trial. Pediatrics. 2006;117:75-83.
22. Bateman DA, Thomas W, Parravicini E, Polesana E, Locatelli C, Lorenz JM. Serum creatinine concentration in very-low-birth-weight infants from birth to 34-36 wk postmenstrual age. Pediatr Res. 2015;77:696-702.
23. JASP. A fresh way to do statistics. [JASP (version 0.95.4) (Computer Software)]. (2025). Available from: https://jasp-stats.org/
24. Onland W, van de Loo M, Offringa M, van Kaam A. Systemic corticosteroid regimens for prevention of bronchopulmonary dysplasia in preterm infants. Cochrane Database Syst Rev. 2023;3:CD010941.
25. Kiss JK, Gajda A, Mari J, Bereczki C. Blood pressure in preterm infants with bronchopulmonary dysplasia in the first three months of life. Pediatr Nephrol. 2024;39:2475-81.
26. Altit G, Bhombal S, Hopper RK, Tacy TA, Feinstein J. Death or resolution: the "natural history" of pulmonary hypertension in bronchopulmonary dysplasia. J Perinatol. 2019;39:415-25.
27. Çakır U, Tayman C. Effect of sex and mode of delivery on systemic inflammatory indices in preterm infants. Cukurova Med J. 2023;48:1347-56.
28. Collaco JM, Aherrera AD, McGrath-Morrow SA. The influence of gender on respiratory outcomes in children with bronchopulmonary dysplasia during the first 3 years of life. Pediatr Pulmonol. 2017;52:217-24.
29. Lee R, Kostina E, Dassios T, Greenough A. Influence of sex on the requirement for and outcomes following late postnatal corticosteroid treatment. Eur J Pediatr. 2023;182:1417-23.
30. Mataloun MMGB, Gibelli MAC, Kato AV, Vaz FAC, Leone CR. Use of corticosteroids and the outcome of infants with bronchopulmonary dysplasia. Revista do Hospital das Clínicas. 1999;54:175-80.
31. Sung TJ. Bronchopulmonary dysplasia: how can we improve its outcomes? Korean J Pediatr. 2019;62:367-73.
32. Kiss JK, Gajda A, Mari J, Nemeth J, Bereczki C. Oscillometric arterial blood pressure in haemodynamically stable neonates in the first 2 weeks of life. Pediatr Nephrol. 2023;38:3369-78.

DOI: 10.4274/anatoljmed.2026.20092
Anatol J Gen Med Res 2026;36(2):222-9

CLINICAL RESEARCH / KLİNİK ARAŞTIRMA

ANATOLIAN JOURNAL of GENERAL MEDICAL RESEARCH

Predictive Role of Tumor Location and Baseline PET-CT Metabolic Parameters in Rectal Cancer Patients Undergoing Neoadjuvant Treatment

Rektum Kanserli Olgularda Primer Tümör Yerleşim Yeri ve Tedavi Öncesi PET-CT Metabolik Parametrelerinin Neoadjuvan Tedavi Sonrası Prognostik Değeri

Murtaza Parvizi[1], Feray Aras[2], Engin Kut[3], Olcay Ak Nalbant[4]

[1]İzmir Tınaztepe University Private Galen Hospital, Department of Radiation Oncology, İzmir, Türkiye
[2]Manisa Celal Bayar University Faculty of Medicine, Department of Nuclear Medicine, Manisa, Türkiye
[3]Manisa City Hospital, Clinic of Medical Oncology, Manisa, Türkiye
[4]Manisa City Hospital, Clinic of Medical Pathology, Manisa, Türkiye

Cite as: Parvizi M, Aras F, Kut E, Ak Nalbant O. Predictive role of tumor location and baseline PET-CT metabolic parameters in rectal cancer patients undergoing neoadjuvant treatment. Anatol J Gen Med Res. 2026;36(2):222-9

Abstract

Objective: This study evaluated the prognostic significance of primary tumor location and baseline ^{18}F-fluorodeoxyglucose positron emission tomography-computed tomography (PET-CT) metabolic parameters in patients with locally advanced rectal cancer treated with neoadjuvant chemoradiotherapy/radiotherapy.

Methods: A total of 168 patients treated between October 2010 and November 2023 were retrospectively analyzed. Baseline maximum standardized uptake value (SUV_{max}), metabolic tumor volume (MTV), total lesion glycolysis (TLG), and tumor location were evaluated for associations with overall survival (OS) and disease-free survival (DFS).

Results: The median age was 61.5 years and the median follow-up was 72.4 months. The 5-year OS and DFS rates were 72.5% and 57.4%, respectively. Tumor location did not correlate significantly with SUV_{max}, MTV, or TLG (p>0.05). Local recurrence occurred in 11 patients (6.5%), predominantly among patients with lower rectal tumors. Tumor location (r=0.161, p=0.042) and SUV_{max} (r=0.289, p=0.025) were significantly associated with local recurrence. In univariate analyses, SUV_{max}, MTV, and TLG were significant prognostic factors for OS and DFS. In multivariate analyses, MTV was an independent prognostic factor for OS [hazard ratio (HR)=1.040, p=0.03], while both SUV_{max} (HR=1.033, p=0.035) and MTV (HR=1.031, p=0.012) were independent prognostic factors for DFS.

Conclusion: Baseline PET-CT parameters, particularly SUV_{max} and MTV, provide significant prognostic information in locally advanced rectal cancer. The higher local recurrence rate in lower rectal tumors indicates a need for closer surveillance.

Keywords: Rectal cancer, neoadjuvant chemoradiotherapy, neoadjuvant radiotherapy, PET-CT

Address for Correspondence/Yazışma Adresi: Asst. Prof., Murtaza Parvizi, İzmir Tınaztepe University Private Galen Hospital, Department of Radiation Oncology, İzmir, Türkiye
E-mail: drparvizi@yahoo.com
ORCID ID: orcid.org/0000-0002-0280-7321

Received/Geliş tarihi: 22.06.2026
Accepted/Kabul tarihi: 30.07.2026
Published date/Yayınlanma tarihi: 31.08.2026

Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of University of Health Sciences Türkiye, İzmir Tepecik Education and Research Hospital. This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

Öz

Amaç: Bu çalışma, neoadjuvan kemoradyoterapi/radyoterapi ile tedavi edilen lokal ileri evre rektum kanserli hastalarda primer tümör yerleşiminin ve başlangıç flor-18 florodeoksiglukoz pozitron emisyon tomografisi-bilgisayarlı tomografi (PET-CT) metabolik parametrelerinin prognostik önemini değerlendirmiştir.

Yöntem: Ekim 2010 ile Kasım 2023 tarihleri arasında tedavi edilen toplam 168 hasta retrospektif olarak analiz edildi. Başlangıç maksimum standartlaştırılmış tutulum değeri (SUV_{max}), metabolik tümör hacmi (MTV), toplam lezyon glikolizi (TLG) ve tümör yerleşiminin genel sağkalım (OS) ve hastalıksız sağkalım (DFS) ile olan ilişkileri değerlendirildi.

Bulgular: Ortanca yaş 61,5; ortanca takip süresi 72,4 ay olarak saptandı. 5 yıllık OS ve DFS oranları sırasıyla %72,5 ve %57,4 idi. Tümör yerleşimi ile SUV_{max}, MTV veya TLG arasında anlamlı bir ilişki bulunmadı (p>0,05). Çoğunluğu alt rektum tümörlerinde olmak üzere 11 hastada (%6,5) lokal (bölgesel) nüks gelişti. Tümör yerleşimi (r=0,161, p=0,042) ve SUV_{max} (r=0,289, p=0,025) lokal nüks ile anlamlı derecede ilişkiliydi. Tek değişkenli analizlerde SUV_{max}, MTV ve TLG; OS ve DFS için anlamlı prognostik faktörler olarak bulundu. Çok değişkenli analizlerde ise MTV, OS için bağımsız bir prognostik faktör iken [tehlike oranı (HR)=1,040, p=0,03]; hem SUV_{max} (HR=1,033, p=0,035) hem de MTV (HR=1,031, p=0,012) hastalıksız sağkalım için bağımsız prognostik faktörler olarak belirlendi.

Sonuç: Başlangıç PET-CT parametreleri, özellikle de SUV_{max} ve MTV, lokal ileri evre rektum kanserinde önemli prognostik bilgiler sağlamaktadır. Alt rektum tümörlerinde daha yüksek lokal nüks oranı, bu hastaların daha yakın takip edilmesi gerektiğini göstermektedir.

Anahtar Kelimeler: Rektum kanseri, neoadjuvan kemoradyoterapi, neoadjuvan radyoterapi, PET-CT

Introduction

Colorectal cancer (CRC) continues to represent one of the major public health challenges worldwide and remains one of the leading causes of cancer-related morbidity and mortality. According to the most recent GLOBOCAN estimates, CRC is the third most common malignancy and the second leading cause of cancer-related death globally, with nearly 1.9 million newly diagnosed cases annually[1]. Ongoing changes in lifestyle patterns, including increasing rates of obesity, sedentary behavior, alcohol consumption, and dietary habits characterized by a high intake of processed foods, are expected to further increase the incidence of CRC in the coming decades[2]. Rectal cancer accounts for approximately one-third of all CRC malignancies and exhibits distinct biological and clinical characteristics compared with colon cancer. Despite significant improvements in multidisciplinary treatment approaches, a considerable proportion of patients are metastatic at diagnosis, whereas others subsequently develop distant metastases following curative-intent treatment[3]. Consequently, identification of prognostic factors that may predict treatment outcomes remains an important area of clinical research.

Several patient-related and tumor-related variables have been associated with prognosis in rectal cancer. Age, sex, comorbid conditions, tumor stage, pathological characteristics, primary tumor location, and treatment modality have all been reported to influence survival and disease control[4-6]. Therefore, comprehensive pretreatment evaluation is essential for appropriate therapeutic decision-making and individualized risk assessment. Current staging algorithms generally rely on a combination of imaging modalities, including contrast-enhanced computed tomography (CT), pelvic magnetic resonance imaging (MRI), and in selected cases, positron emission tomography-CT (PET-CT)[7-9].

Among contemporary techniques, ^{18}F-fluorodeoxyglucose (^{18}F-FDG) PET-CT provides functional information regarding tumor metabolism in addition to anatomical assessment. PET-CT demonstrates high sensitivity for the detection of the primary tumor location. Previous studies have demonstrated excellent performance in identifying primary colorectal tumors and distant metastatic lesions, although its performance in evaluating regional lymph node involvement remains limited compared with MRI-based assessment[10-12]. Beyond its diagnostic utility, PET-CT enables quantification of several metabolic parameters, including maximum standardized uptake value (SUV_{max}), metabolic tumor volume (MTV), and total lesion glycolysis (TLG)[9]. These parameters have attracted considerable interest because they may reflect tumor aggressiveness and biological behavior. Previous studies have suggested that high metabolic activity at diagnosis may be associated with worse oncological outcomes in patients undergoing treatment for rectal cancer[13].

Primary tumor location has also emerged as a clinically relevant prognostic factor in rectal cancer. Anatomically, the rectum is commonly subdivided into upper, middle and lower segments based on the distance from the anal

verge[14]. Variations in pelvic anatomy, lymphatic drainage pathways, vascular supply, and surgical accessibility may result in different treatment responses and clinical outcomes for tumors arising from these regions. In particular, tumors located in the lower rectum have frequently been associated with higher local recurrence rates and greater technical challenges during surgical management compared with tumors arising in more proximal rectal segments[6,7,14].

The current standard of care for locally advanced rectal cancer (LARC) consists of neoadjuvant chemoradiotherapy/radiotherapy (nCRT/nRT) followed by total mesorectal excision, an approach that has been shown to enhance local tumor control and improve resectability[5]. Although several investigations have reported the prognostic relevance of PET-CT metabolic parameters and tumor location separately, studies assessing the interplay between these variables and their influence on treatment outcomes remain limited. In this study, we investigated the association of baseline ^{18}-FDG PET-CT metabolic parameters and primary tumor localization with survival and treatment outcomes in patients with LARC treated with nCRT/nRT.

Materials and Methods

Study Type and Ethics

Patients diagnosed with LARC who underwent nCRT/nRT at the Radiation Oncology Clinic of Manisa City Hospital between October 2010 and November 2023 were included in this retrospective cohort study. The research protocol underwent ethical review and was approved by the Ethics Committee of Manisa Celal Bayar University (approval no: 20.478.486/3338, date: 30.07.2025), and by the administration of Manisa City Hospital under approval number E-67427165-605-279739188.

Patient Selection

The medical records of 194 consecutive patients who underwent treatment for LARC during the study period were retrospectively reviewed. All patients were treatment-naïve at the time of initial evaluation and had been referred for neoadjuvant treatment as part of a multidisciplinary management strategy.

The inclusion criteria included age ≥18 years, histopathological confirmation of rectal adenocarcinoma, availability of complete clinical, radiological, pathological, treatment-related data, and a minimum follow-up duration of 12 months.

The exclusion criteria included: age <18 years, presence of another primary malignancy, unavailability of complete clinical and treatment data, non-adenocarcinoma histology, or follow-up duration less than 12 months.

Data Collection

Patient information was obtained from the institutional electronic database and the archived medical records. Demographic characteristics, presenting symptoms, comorbid diseases, smoking and alcohol history, family history of cancer, colonoscopy findings, histopathological reports, radiological imaging results, treatment details, and follow-up data were collected. Tumor localization was determined from colonoscopy findings and categorized, based on distance from the anal verge, as lower rectum (0-5 cm), middle rectum (5.1-10 cm), or upper rectum (>10 cm). Baseline staging investigations, surgical reports, pathological findings, recurrence patterns, metastatic progression, survival status, and dates of follow-up were also recorded for analysis.

PET-CT Image Analysis

Baseline PET-CT images were retrospectively reviewed using the GE Advantage Workstation. Metabolic parameters, including SUV_{max}, MTV, and TLG, were recorded. Tumor segmentation was performed automatically using a standardized threshold of 42% of SUV_{max}, and MTV and TLG were calculated based on this segmentation method.

Statistical Analysis

Statistical evaluation was performed using specialized software. Continuous data were summarized according to their distribution characteristics and reported as mean ± standard deviation or median (range), while categorical variables were expressed as frequencies and percentages. Relationships between clinical variables and PET-CT metabolic parameters were assessed using Pearson or Spearman correlation analyses, depending on the distribution of the variables. Variables identified as statistically significant in univariate analyses were subsequently included in multivariate Cox proportional hazards regression models to determine independent prognostic factors. A two-sided p-value of less than 0.05 was considered statistically significant. Survival outcomes were estimated using the Kaplan-Meier method, and 5-year survival probabilities were calculated with life-table analyses.

Results

Based on the study's inclusion and exclusion criteria, the medical records and clinical data of 194 patients were reviewed, of whom 26 were excluded from the analysis. Consequently, data from a total of 168 patients were included in the final evaluation. Among these patients, 68 (40.5%) were women and 100 (59.5%) were men, with a median age of 61.5 years (range, 28-85 years).

As shown in Table 1, the most common presenting symptom was rectal bleeding, observed in 64.3% of patients. Evaluation of comorbidities revealed that diabetes mellitus and cardiovascular diseases were the most frequent accompanying conditions, while 38.1% of the patients were smokers and 8.3% reported alcohol consumption. A family history of cancer was present in 11.4% of patients, with gastrointestinal system malignancies being the most common cancer type (5.4%, n=9). Following diagnostic colonoscopy, tumor localization was identified as lower rectum (0-≤5 cm, 50%, n=84), middle rectum (5.1-≤10 cm, 39.9%, n=67), and upper rectum (≥10 cm, 10.1%, n=17), respectively. Histopathological examination of all colonoscopic biopsy specimens revealed adenocarcinoma. Most of the patients (125 patients, 73.5%) had a primary tumor size ≥50 mm at diagnosis. Pre-treatment clinical stages were determined according to initial clinical and radiological staging as follows: stage IIIB in 109 patients (64.9%), stage IIIC in 30 patients (17.9%), stage IIA in 26 patients (15.4%), stage IIC in 2 patients (1.2%), and stage IIB in 1 patient (0.6%).

nCRT/nRT was delivered according to the treatment technique as follows: 35 patients (20.8%) received three-dimensional conformal radiotherapy (45 Gy to the primary tumor and regional lymphatics in 25 fractions), whereas 133 patients (78.2%) underwent intensity-modulated radiotherapy, consisting of 45 Gy to the primary tumor and regional lymphatics with a simultaneous integrated boost to the primary tumor up to 50 Gy in 25 fractions. Concurrent chemotherapy was not administered to 8 patients because of advanced age or comorbidities. Among the 160 patients who received concurrent chemotherapy, 38 patients (29.4%) were treated with 5-fluorouracil/leucovorin (Mayo regimen), whereas 122 patients (65.8%) received oral capecitabine. Of the 143 patients (85.1%) who subsequently underwent surgery following nCRT/nRT, 98 (58.3%) underwent low anterior resection and 45 (26.8%) underwent abdominoperitoneal resection. In the postoperative period, a total of 124 patients received adjuvant chemotherapy, including 13 patients (7.6%) with 5-fluorouracil/leucovorin (Mayo regimen), 13 patients (7.6%) with oral capecitabine, 84 patients (50%) with capecitabine plus oxaliplatin, and 14 patients (8.3%) with FOLFOX regimen.

Correlation analyses of baseline ^{18}F-FDG PET-CT findings related to primary tumor characteristics demonstrated weak negative correlations between primary tumor localization

Table 1. Demographic and clinical characteristics of the patients (n=168)

Parameters		
Age, years [median (min-max)]		61.5 (28-85)
		n (%)
Gender	Female	68 (40.5)
	Male	100 (59.5)
Smoking history	No	104 (61.9)
	<20 pack-years	42 (25)
	≥20 pack-years	22 (13.1)
Alcohol-drinking history	No	154 (91.7)
	Yes	14 (8.3)
First symptom	Rectal bleeding	108 (64.3)
	Constipation	22 (13.1)
	Multiple	17 (10.1)
	Pelvic pain and rectal bleeding	11 (6.5)
	Diarrhea	10 (6)

Table 1. Continued

Parameters		
Additional diseases	No	109 (64.9)
	Multiple	19 (11.3)
	Diabetes mellitus	14 (8.3)
	Cardiovascular disease	12 (7.1)
	Hypertension	9 (5.4)
	Chronic obstructive pulmonary disease	3 (1.8)
	Ulcerative colitis	2 (1.2)
Family history	No	149 (88.6)
	Gastrointestinal cancer	9 (5.4)
	Genitourinary cancer	2 (1.2)
	Breast cancer	2 (1.2)
	Other	6 (3.6)
Tumor location	Lower rectum (1-5≥ cm)	84 (50)
	Mid rectum (5.1-10≥ cm)	67 (39.9)
	Upper rectum (>10 cm)	17 (10.1)
Tumor size	<50 mm	43 (26.5)
	≥50 mm	125 (73.5)
Clinical stage	IIA	26 (15.4)
	IIB	1 (0.6)
	IIC	2 (1.2)
	IIIB	109 (64.9)
	IIIC	30 (17.9)
Type of neoadjuvant treatment	Radiotherapy	8 (4.8)
	Chemoradiotherapy	160 (95.2)
Concurrent chemotherapy	No	8 (4.8)
	5-flourouracil	38 (29.4)
	Oral capecitabine	122 (65.8)
Type of operation	No	25 (14.9)
	LAR	98 (58.3)
	APR	45 (26.8)

LAR: Low anterior resection, APR: Abdominoperineal resection

and SUV_{max} (r=0.147, p=0.249), MTV (r=0.043, p=0.798), and TLG (r=0.033, p=0.685). However, none of these correlation coefficients reached statistical significance (p>0.05).

The median follow-up duration for the patients was 72.4 months (range, 12.0-176.5 months). Local recurrences developed in 11 patients (6.5%) during follow-up, with a median time to recurrence of 23.75 months (range, 12.5-45.5 months). Additionally, distant metastases developed in 42 patients (24%); of these, 19 patients (11.3%) had liver metastases, with a median time from diagnosis to metastasis of 23.5 months (range, 13.5-45.5 months). Overall survival (OS) was 122 months (71.5-176.5 months), and the 5-year OS rate was 72.5%. Disease-free survival (DFS) was 75.0 months (range, 42.5-176.5 months), and the 5-year DFS rate was 57.4%.

Evaluation of the 11 patients who developed local recurrence following therapy demonstrated that recurrences predominantly occurred in the lower rectum (7 patients). A weak but statistically significant positive correlation was observed between tumor localization and local recurrence

(r=0.161, p=0.042). Regarding diagnostic ^{18}F-FDG PET-CT parameters in patients with local recurrence, SUV_{max} showed a statistically significant association (r=0.289, p=0.025), whereas no significant correlation was observed for the MTV and TLG parameters (r=0.043, p=0.07 and r=0.064, p=0.63, respectively).

We evaluated the surgical margins in patients who developed local recurrence. The median closest distal/proximal surgical margin was 38 mm (range, 24-50 mm), and the median radial surgical margin was 7 mm (range, 1.0-10 mm). In patients without recurrence, the median closest distal or proximal surgical margin was 35 mm (range, 20-80 mm), and the median radial margin was 9 mm (range, 1.0-16 mm). No statistically significant correlation was identified between surgical margins and local recurrence.

The prognostic significance of baseline staging ^{18}F-FDG PET-CT parameters in rectal cancer patients treated with nCRT/nRT was evaluated and presented in Table 2. In univariate analysis, SUV_{max} [hazard ratio (HR): 1.018, 95% confidence interval (CI): 1.010-1.040, p=0.032], MTV (HR: 1.032, 95% CI: 1.007-1.057, p=0.011), and TLG (HR: 1.025, 95% CI: 1.015-1.030, p=0.041) were identified as significant prognostic factors for OS. However, in multivariate analysis, only MTV remained an independent prognostic factor (HR: 1.040, 95% CI: 1.004-1.078, p=0.03). Similarly, DFS, SUV_{max}, MTV, and TLG parameters were statistically significant in univariate analysis (p=0.002, p=0.001, and p=0.003). In multivariate analysis, SUV_{max} (HR: 1.033, 95% CI: 1.002-1.064, p=0.035) and MTV (HR: 1.031, 95% CI: 1.007-1.057, p=0.012) remained independent prognostic factors, whereas TLG was no longer statistically significant (p=0.33).

Discussion

Accurate prognostic assessment following nCRT/nRT remains a critical component of contemporary management strategies for patients with LARC. Although numerous clinicopathological factors have been associated with treatment outcomes, increasing interest has focused on the prognostic role of metabolic parameters obtained from baseline ^{18}F-FDG PET-CT imaging[13,15]. In the present study, we evaluated the association of primary tumor location and baseline PET-derived metabolic parameters with oncological outcomes following nCRT/nRT. This study demonstrated that lower rectal tumors were associated with a higher risk of local recurrence and that baseline SUV_{max}, MTV, and TLG values showed prognostic relevance. Notably, no significant association was identified between primary tumor location and PET-derived metabolic parameters.

The prognostic implications of primary tumor location in rectal cancer have been investigated extensively. Tumor arising in the lower rectum are often associated with increased surgical complexity because of anatomical constraints within the pelvis, challenges in achieving adequate circumferential resection margins, and limitations related to sphincter-preserving procedures[16]. Consistent with previous reports, patients with lower rectal tumors in our cohort experienced a higher incidence of local recurrence. Surgical margin status was not significantly associated with local recurrence. This finding may reflect the relatively limited number of recurrence events in our study population and may suggest that factors beyond surgical technique alone, including tumor biology and treatment response, may contribute to disease relapse. Another important observation in our study was the absence of a significant relationship between primary tumor location and baseline metabolic activity, measured by PET-CT. Although tumor located in different rectal segments exhibit distinct anatomical characteristics and clinical outcomes, these differences were not accompanied by measurable variations in SUV_{max}, MTV, or TLG values[6,7,14]. This finding suggests that the poorer outcomes observed in lower rectal tumors may be driven primarily by anatomical and treatment-related factors rather than differences in intrinsic metabolic aggressiveness. Consequently, primary

Table 2. Overall survival (OS) and disease-free survival (DFS) of patients according primer PET-CT finding

	OS				DFS			
PET-CT finding	**Univariate analysis (HR, 95% CI)**	**p-value**	**Multivariate analysis (HR, 95% CI)**	**p-value**	**Univariate analysis (HR, 95% CI)**	**p-value**	**Multivariate analysis (HR, 95% CI)**	**p-value**
SUV_{max}	1.018 (1.010-1.040)	**0.032**	1.002 (0.999-1.004)	0.47	1.019 (1.010-1.043)	**0.002**	1.033 (1.002-1.064)	**0.035**
MTV	1.032 (1.007-1.057)	**0.011**	1.040 (1.004-1.078)	**0.03**	1.046 (1.019-1.0739	**0.001**	1.031 (1.007-1.057)	**0.012**
TLG	1.025 (1.015-1.030)	**0.041**	1.027 (0.966-1.070)	0.53	1.007 (1.002-1.012)	**0.003**	1.027 (1.003-1.046)	0.33

PET-CT: Positron emission tomography-computed tomography, SUV_{max}: Maximum standardized uptake value, MTV: Metabolic tumor volume, TLG: Total lesion glycolysis, CI: Confidence interval, HR: Hazard ratio

tumor location and PET-derived metabolic parameters may provide complementary prognostic information rather than reflecting the same biological processes.

Among the evaluated metabolic parameters, SUV_{max} was significantly associated with local recurrence. High FDG uptake has been linked to enhanced glucose metabolism, accelerated cellular proliferation, tumor hypoxia, and resistance to anticancer therapies, all of which may contribute to more aggressive tumor behavior[15,17]. Therefore, evaluated baseline SUV_{max} may identify tumors that are more likely to recur despite multimodality treatment. However, the prognostic significance of SUV_{max} remains controversial. Deantonio et al.[18] reported that baseline SUV_{max} was not independently associated with DFS or OS, whereas MTV showed stronger correlation with treatment response. Similarly, Fernando et al.[19] demonstrated that serial PET-CT assessments performed during and after neoadjuvant treatment provided valuable prognostic information, while baseline SUV_{max} alone showed limited predictive value for long-term outcomes. In contrast, other studies have demonstrated significant associations between elevated pre-treatment SUV_{max} values and unfavorable oncological outcomes[13]. Variability in patient characteristics, imaging acquisition protocols, segmentation techniques, and cut-off definitions may explain these inconsistent findings across studies.

One of the most clinically relevant findings of the present study was the identification of MTV as an independent prognostic factor for both OS and DFS. Unlike SUV_{max}, which reflects the highest metabolic activity within a single tumor region, MTV quantifies the total volume of metabolically active tumor tissue and may therefore better represent overall tumor burden. This broader assessment of tumor biology may explain its superior prognostic performance. Similar findings have been reported previously by Choi et al.[15], who demonstrated that MTV independently predicted both survival outcomes in patients with rectal adenocarcinoma. These results support the potential utility of MTV as a clinically important imaging biomarker for pre-treatment risk stratification in LARC.

Although TLG was significantly associated with OS and DFS in univariate analyses, it failed to maintain its independent significance in multivariate models. Since TLG includes both metabolic activity and metabolically active tumor volume, strong correlations with MTV may reduce its independent prognostic contribution when both parameters are evaluated. Similar observations have been reported in previous PET-CT studies investigating rectal cancer outcomes[13,15,18].

Recent advances in radiomics and artificial intelligence have expanded the potential applications of metabolic imaging beyond conventional SUV-based metrics. Texture features extracted from FDG PET-CT images may provide additional information regarding intratumoral heterogeneity and biological behavior, potentially improving prediction of treatment response and survival outcomes[20]. As these technologies continue to evolve, the integration of advanced imaging biomarkers into prognostic models may further enhance personalized treatment approaches for patients with rectal cancer.

Study Limitations

The strengths of our study include a relatively long follow-up duration, a homogeneous patient population, the use of modern radiotherapy techniques, and the combined evaluation of other metabolic parameters, primary tumor localization, and SUV_{max} values using FDG PET-CT. However, several limitations should be acknowledged. First, the retrospective, single-center design of the study represents a major limitation. Additionally, the lack of standardization in the segmentation methods we use to measure FDG PET-CT metabolic parameters may limit the external validity of these methods. The absence of molecular biomarkers and radiomic analyses in the present study should be considered an important limitation. Therefore, further multicenter studies are needed to confirm these findings.

Conclusion

These findings highlight the prognostic relevance of pretreatment ^{18}F-FDG PET-CT metabolic parameters in patients with LARC undergoing nCRT/nRT. Among these parameters, MTV emerged as an independent predictor of OS and DFS, highlighting the potential importance of metabolic tumor burden in pre-treatment risk assessment. Additionally, the increased risk of local recurrence observed in lower rectal tumors supports the need for individualized surveillance strategies in this subgroup. The lack of correlation between the location of the primary tumor and metabolic PET-CT parameters further suggests that anatomical and metabolic tumor characteristics may represent different prognostic domains. Prospective multicenter studies are warranted to validate these findings and clarify their potential role in personalized treatment planning.

Ethics

Ethics Committee Approval: The research protocol underwent ethical review and was approved by the Ethics Committee of Manisa Celal Bayar University (approval no: 20.478.486/3338, date: 30.07.2025), and by the administration of Manisa City Hospital under approval number E-67427165-605-279739188.

Informed Consent: Retrospective design.

Footnotes

Authorship Contributions

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

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

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

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74:229-63.
2. Xi Y, Xu P. Global colorectal cancer burden in 2020 and projections to 2040. Transl Oncol. 2021;14:101174.
3. Cervantes A, Adam R, Roselló S, et al. Electronic address: clinicalguidelines@esmo.org. Metastatic colorectal cancer: ESMO Clinical Practice Guideline for diagnosis, treatment and follow-up. Ann Oncol. 2023;34:10-32.
4. Khalfallah M, Dougaz W, Jerraya H, Nouira R, Bouasker I, Dziri C. Prognostic factors in rectal cancer: where is the evidence? Tunis Med. 2017;95:79-86.
5. Li Y, Wang J, Ma X, et al. A review of neoadjuvant chemoradiotherapy for locally advanced rectal cancer. Int J Biol Sci. 2016;12:1022-31.
6. Wang CB, Shahjehan F, Merchea A, et al. Impact of tumor location and variables associated with overall survival in patients with colorectal cancer: a mayo clinic colon and rectal cancer registry study. Front Oncol. 2019;9:76.
7. Lee MH, Kim DH. Low rectal cancers at initial staging MRI. Radiographics. 2023;43:e230080.
8. Azin A, Wood T, Hirpara D, et al. Diagnostic utility of staging abdominal computerized tomography and repeat endoscopy in detecting localization errors at initial endoscopy in colorectal cancer. Surg Endosc. 2018;32:3303-10.
9. Heo SH, Kim JW, Shin SS, Jeong YY, Kang HK. Multimodal imaging evaluation in staging of rectal cancer. World J Gastroenterol. 2014;20:4244-55.
10. Ye Y, Liu T, Lu L, et al. Pre-operative TNM staging of primary colorectal cancer by (18)F-FDG PET-CT or PET: a meta-analysis including 2283 patients. Int J Clin Exp Med. 2015;8:21773-85.
11. Lu YY, Chen JH, Ding HJ, Chien CR, Lin WY, Kao CH. A systematic review and meta-analysis of pretherapeutic lymph node staging of colorectal cancer by 18F-FDG PET or PET/CT. Nucl Med Commun. 2012;33:1127-33.
12. Kwak JY, Kim JS, Kim HJ, Ha HK, Yu CS, Kim JC. Diagnostic value of FDG-PET/CT for lymph node metastasis of colorectal cancer. World J Surg. 2012;36:1898-905.
13. Li J, Zhou Y, Liu L, Pang H. Diagnostic and prognostic value of quantitative ^{18}F-FDG PET/CT metabolic parameters combined with clinical indicators in patients with locally recurrent rectal cancer. Abdom Radiol (NY). 2025;50:5135-47.
14. Wang YHW, Wiseman J. Anatomy, abdomen and pelvis, rectum. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026.
15. Choi BW, Kang S, Bae SU, et al. Prognostic value of metabolic parameters on ^{18}F-fluorodeoxyglucose positron tomography/computed tomography in classical rectal adenocarcinoma. Sci Rep. 2021;11:12947.
16. Dijkstra EA, Hospers GAP, Kranenbarg EM, et al. Quality of life and late toxicity after short-course radiotherapy followed by chemotherapy or chemoradiotherapy for locally advanced rectal cancer - the RAPIDO trial. Radiother Oncol. 2022;171:69-76.
17. Hanahan D. Hallmarks of cancer: new dimensions. Cancer Discov. 2022;12:31-46.
18. Deantonio L, Caroli A, Puta E, et al. Does baseline [18F] FDG-PET/CT correlate with tumor staging, response after neoadjuvant chemoradiotherapy, and prognosis in patients with rectal cancer? Radiat Oncol. 2018;13:211.
19. Fernando S, Lin M, Pham TT, et al. Prognostic utility of serial ^{18}F-FDG-PET/CT in patients with locally advanced rectal cancer who underwent tri-modality treatment. Br J Radiol. 2020;93:20190455.
20. Ju HM, Yang J, Park JM, et al. Prediction of neoadjuvant chemoradiotherapy response in rectal cancer patients using harmonized radiomics of multcenter ^{18}F-FDG-PET image. Cancers (Basel). 2023;15:5662.

DOI: 10.4274/anatoljmed.2026.e001
Anatol J Gen Med Res 2026;36(2):230

ERRATUM

OPEN ACCESS

DOI: 10.4274/anatoljmed.2025.19870

Yıldırım Ş, İnci A, Gül S, Yılmaz Ü, Aslaner A. The prognostic significance of Banff classification lesions in determining renal allograft survival during chronic active antibody-mediated rejection. Anatol J Gen Med Res. 2026;36(1):21-8.

The footnotes of Figure 2 have been revised. The incorrect version is presented below:

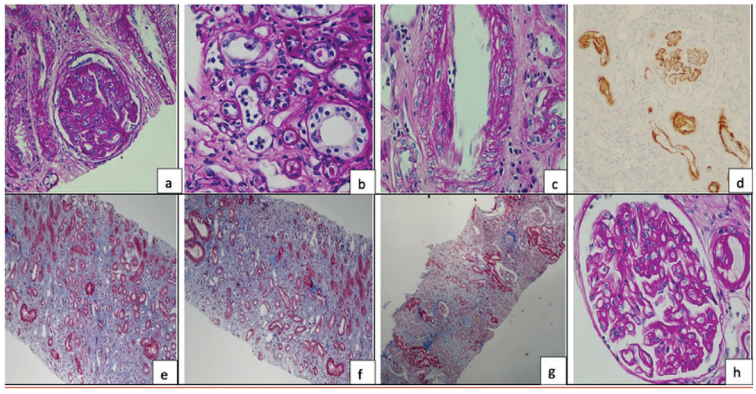
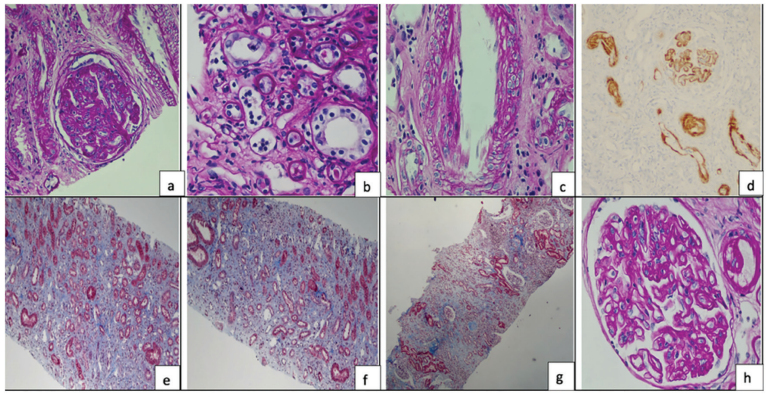


Figure 2. Acute (a,b,c,d) and chronic (e,f,g,h) Banff lesions in biopsies diagnosed with ca-ABMR

a: Glomerulitis (g) (PAS, x200), b: Peritubular capillaritis (pct) (PAS, x200) c: Intimal arteritis (v) (HE, x400) d: Positive C4d staining along the capillary basement membranes of glomeruli and in arterioles (C4d, x100), e: Interstitial fibrosis (ci) (Masson trichrome, x100) f: Tubular atrophy (ct) (Masson trichrome, x100), g: Total inflammation (ti) (Masson trichrome, x40), h: Transplant glomerulopathy (cg) (PAS, x400), ca-ABMR: Chronic active antibody-mediated rejection, PAS: Periodic acid-schiff, HE: Hematoxylin and eosin



An additional note has been added to the footnotes of Figure 2, which have been revised as follows:

Figure 2. Acute (a,b,c,d) and chronic (e,f,g,h) Banff lesions in biopsies diagnosed with ca-ABMR

a: Glomerulitis (g) (PAS, x200), b: Peritubular capillaritis (pct) (PAS, x200) c: Intimal arteritis (v) (HE, x400) d: Positive C4d staining along the capillary basement membranes of glomeruli and in arterioles (C4d, x100), e: Interstitial fibrosis (ci) (Masson trichrome, x100) f: Tubular atrophy (ct) (Masson trichrome, x100), g: Total inflammation (ti) (Masson trichrome, x40), h: Transplant glomerulopathy (cg) (PAS, x400), ca-ABMR: Chronic active antibody-mediated rejection, PAS: Periodic acid-schiff, HE: Hematoxylin and eosin

(Figures 2a, 2b, 2c, and 2e are retrieved from our previous study titled "Demographic and Histopathological Features of Graft Dysfunction in Renal Transplant Biopsies: A Retrospective Study" published in Transplantation Proceedings (https://doi.org/10.1016/j.transproceed.2025.07.007), and are used here in accordance with the author rights policy provided by the publisher, Elsevier)