ISSN 2651-3153 • EISSN 2651-3153

European Archives of Medical Research

Formerly Okmeydanı Medical Journal

Volume: 42 • Number: 2 • June 2026

Impact of Deep Learning–Based Real-Time Polyp Detection on Adenoma Detection Rate and Optical Biopsy Accuracy in Routine Colonoscopy

Prospective pragmatic study · Turkey · Oct 2025 – Feb 2026 · n = 2,122 colonoscopies

Highlights

- Primary and Secondary Autoimmune Hemolytic Anemia
  Fatma Pinar Acar, Naciye Demirel, Caner Acar, Pelin Kerter, Mehmet Kucuk
- Optimizing Medial Release in Varus TKA
  Serhat Gurbuz, Mustafa Bugra Ayaz, Niyazi Igde, Hakan Bahar, Ahmet Keskin, Hakan Gurbuz
- Mini-PNL and RIRS in Children
  Recep Burak Degirmentepe, Yasir Muhammed Akca, Deniz Gul, Haci Ibrahim Cimen, Muammer Bozkurt, Tuncay Toprak, Fikret Halis
- Nurses' Knowledge and Attitude Towards Nutrition
  Nese Iscan Ayyildiz, Sevgi Demir Cam
- Gait Outcomes After SEMLS in CP
  Kadir Cevik, Dilan Deniz Yilmaz, Zeynep Altinkaya, Didem Derici Yildirim, Ugur Dal, Abtullah Milcan
- BI-RADS Classification: Correlation and Pitfalls
  Selçuk Cin, Gül Gizem Pamuk, Muhammed Edip İşleyen, Özgecan Gündoğar, Eda Cingöz, Murat Yüce, Merve Cin, Ali Muhammedoğlu, Ahmet Tan Cimilli
- Comparison of HALP and Ranson Scores in Acute Pancreatitis
  Ibrahim Altundag, Alperen Sahin, Metin Oner, Kaan Yusufoglu, Burcu Genc Yavuz, Sinem Dogruyol, Mustafa Ahmet Afacan, Sahin Colak
- Breast Cancer and Neoadjuvant Pertuzumab
  Orcun Can, Seval Ay, Muhammed Mustafa Atci
- Comparison English vs Turkish Clubfoot Videos
  Tahsin Olgun Bayraktar, Osman Sariyar, Serdar Aki, Nazım Erkurt, Mehmet Selcuk Saygili, Hakan Gurbuz
- Fatigue in Pregnant Women
  Elif Ucar, Melih Bestel
- Chronic Pain and Nurse Attitudes
  Özge Demir, Ezgi Demirturk Selcuk
- Screen Exposure in Enuresis
  Ozgur Ozdemir Simsek, Gamze Yakin, Hasan Turan, Hasan Akin Tahillioglu
- Ultrasound Radiomics for Skin Lesion Classification
  Hakan Ayyildiz
- AI-Assisted Colonoscopy and Adenoma Detection
  Yavuz Ozden, Ferhat Omurca, Dilek Tekis, Oktay Bulur, Nuh Mehmet Buyukberber
- PCOS Risk After Premature Adrenarche
  Didem Helvacioglu, Aylin Tugba Canbaz, Busra Gurpinar Tosun, Tulay Guran

KARE MEDYA

Official Journal of University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital
eurarchmedres.org

European Archives of Medical Research

Formerly Okmeydanı Medical Journal

Owner on behalf and Responsible Manager of University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital

Ali Alemdar

Clinic of General Surgery, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye

Editor in Chief

İsmail Demirkale

Clinic of Orthopedics and Traumatology, University of Health Sciences Türkiye, Sisli Hamidiye Etfal Training and Research Hospital, istanbul, Türkiye

ORCID ID: 0000-0001-7230-1599

Associate Editors

Asım Kalkan

Clinic of Emergency Medicine, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye

ORCID: orcid.org/0000-0002-5800-0201

Müjdat Adaş

Clinic of Orthopedics and Traumatology, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye

ORCID: orcid.org/0000-0003-3637-8876

Tamer Özülker

Clinic of Nuclear Medicine, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye

ORCID ID: 0000-0001-9521-683X

Namigar Turgut

Clinic of Anesthesia and Reanimation, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye

ORCID: orcid.org/ 0000-0003-0252-3377

Özben Yalçın

Clinic of Pathology, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye

ORCID: orcid.org/0000-0002-0019-1922

Editorial Staff

Eda Ertuğrul Usta

University of Health Sciences Turkey, Prof. Dr. Cemil Taşcıoğlu City Hospital, Department of Strategy and Press Office, İstanbul, Turkey

E-mail: stratejicemiltascioglu@gmail.com

ORCID ID: 0000-0003-4427-5332

Editorial Board

Achmet Ali

Department of Anesthesiology and Reanimation, İstanbul University Faculty of Medicine, İstanbul, Türkiye

ORCID ID: 0000-0002-7224-6654

Ali Cahid Civelek

Clinic of Radiology, Division of Nuclear Medicine, Johns Hopkins Medical Instituons, Baltimore, USA

ORCID ID: 0000-0003-4637-6292

Alper Ötünçtemur

Clinic of Urology, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye

ORCID ID: 0000-0002-0553-3012

Andrej Nikolovski

Department of Visceral Surgery, University Surgery Clinic Sv. Naum Ohridski; Ss. Cyril and Methodius University in Skopje, Skopje, Republic of North Macedonia

ORCID ID: 0000-0002-5286-3532

Arzu Akan

Clinic of General Surgery, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye

ORCID ID: 0000-0001-8435-9771

Berin Upcin

Institute of Anatomy and Cell Biology, Julius-Maximilians-University, Würzburg, Germany

ORCID ID: 0000-0003-4853-9358

Berrin Hüner

Clinic of Physical Therapy and Rehabilitation, Gaziosmanpaşa Training and Rearch Hospital, İstanbul, Türkiye

ORCID ID: 0000-0003-3584-8880

Burak Erden

Clinic of Eye Diseases, Dünya Göz Hospital Ataköy, İstanbul, Türkiye

ORCID ID: 0000-0003-0650-4552

Bülent Ozgonenel

Clinic of Hematology Oncology, Children's Hospital of Michigan, Detroit, United States

ORCID ID: 0000-0001-8891-7646

Ekrem Üçer

University Hospital Regensburg, Department of Cardiology, Regensburg, Germany

ORCID ID: 0000-0002-3935-1110

Funda Şimşek

Clinic of Infectious Diseases and Departmental Microbiology, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye

ORCID ID: 0000-0002-7387-5057

Gülcan Güntaş

Department of Biochemistry, Faculty of Medicine, Kırklareli University, Kırklareli Türkiye

ORCID ID: 0000-0002-3638-4662

Hakan Önder

Clinic of Radiology, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye

ORCID ID: 0000-0001-5207-3314

Hasan Dursun

Clinic of Pediatrics, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye

ORCID ID: 0000-0002-8817-494X

European Archives of Medical Research

Formerly Okmeydanı Medical Journal

İlteriş Oğuz Topal
Clinic of Dermatology, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye
ORCID ID: 0000-0001-8735-9806
Kadriye Kılıçkesmez
Clinic of Cardiology, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye
ORCID ID: 0000-0002-2139-9909
Nurdan Gül
Department of Endocrinology, İstanbul University Faculty of Medicine, İstanbul, Türkiye
ORCID ID: 0000-0002-1187-944X
Mehmet Küçük
Clinic of Internal Medicine, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye
ORCID ID: 0000-0003-1720-3819
Mete Gürsoy
Clinic of Cardiovascular Surgery, University of Health Sciences Türkiye, Mehmet Akif Ersoy Chest and Cardiovascular Surgery Hospital, İstanbul, Türkiye
ORCID ID: 0000-0002-7083-476X
Metin Çetiner
Duisburg-essen University School Of Medicine, Division Of Pediatric Nephrology And Pediatric Sonography Hufelandstrate Ss
ORCID ID: 0000-0002-0918-9204
Mine Adaş
Clinic of Internal Medicine, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye
ORCID ID: 0000-0003-3008-6581
Murat Dursun
Department of Urology, İstanbul University Faculty of Medicine, İstanbul, Türkiye
ORCID ID: 0000-0001-9115-7203
Özge Kandemir Gürsel
Clinic of Radiation Oncology, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye
ORCID ID: 0000-0002-6960-4115
Seçil Arıca
Clinic of Family Practice, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye
ORCID ID: 0000-0003-0135-6909
Seçkin Aydın
Clinic of Brain Surgery, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye
ORCID ID: 0000-0003-1195-9084
Serdar Günaydın
Clinic of Cardiovascular Surgery, University of Health Sciences Türkiye, Ankara City Hospital, Ankara, Türkiye
ORCID ID: 0000-0002-9717-9793
Sezen Karakuş
Department of Ophthalmology, The Johns Hopkins Wilmer Eye Institute, Baltimore, USA
ORCID ID: 0000-0003-2951-995X
Sinan Akay
Department of Radiology, University of Iowa Health Care, 52242, Iowa City, IA
ORCID ID: 0000-0001-7201-475X
Şener Cihan
Department of Medical Oncology, İstinye University Faculty of Medicine, Medical Park Gaziosmanpaşa Hospital, İstanbul, Türkiye
ORCID ID: 0000-0002-3594-3661
Şerife Şimşek
Department of General Surgeon, Breast Surgeon, Fakeeh University Hospital, Dubai, UAE
ORCID ID: 0000-0003-0463-2710
Tolgar Lütfi Kumral
Clinic of Otorhinolaryngology, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye
ORCID ID: 0000-0001-8760-7216
Veli Mihmanlı
Clinic of Gynecology and Obstetrics, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye
ORCID ID: 0000-0001-8701-8462
Yavuz Anacak
Department of Radiation Oncology, Ege University, İzmir, Türkiye
ORCID ID: 0000-0002-2548-1109
Yavuz Uyar
Clinic of Otorhinolaryngology, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye
ORCID ID: 0000-0001-8732-4208
Yücel Arman
Clinic of Internal Medicine, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye
ORCID ID: 0000-0002-9584-6644
Ziya Akçetin
KMG Klinikum Urology Clinic Chief, Luckenwalde, Germany

Statistics Editor

Zübeyde Arat
E-mail: zubeyde@aratistatistik.com
ORCID ID: 0009-0008-1751-686X

Social Media Editor

Caner Baran
Clinic of Urology, University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital, İstanbul, Türkiye
E-mail: drcanerbaran@hotmail.com
ORCID ID: 0000-0002-6315-6518

Publisher Contact

	Prof. Dr. Cemil Taşcıoğlu City Hospital
Address	*University of Health Sciences Türkiye, Prof. Dr. Cemil Taşcıoğlu City Hospital Clinical of E.N.T. / Okmeydani / Istanbul*
E-mail	*info@okmeydanitipdergisi.org*
Website	*cemiltascioglush.saglik.gov.tr*

Publishing House

Kare Publishing

Address	*Göztepe Mah. Fahrettin Kerim Gökay Caddesi No: 200/A D:2 Çemenzar - Kadıköy, İstanbul-Türkiye*
Phone	*+90 216 550 61 11*
Fax	*+90 216 550 61 12*
Web	*www.karepb.com*
E-mail	*kare@karepb.com*

**For requests concerning advertising on printed versions and on the webpage of European Archives of Medical Research, please contact the publisher KARE Publishing.*

European Archives of Medical Research

Formerly Okmeydanı Medical Journal

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

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

European Archives of Medical Research is currently indexed in **TUBITAK ULAKBIM TR Index, Gale, ProQuest, Türk Medline, Türkiye Atıf Dizini, J-GATE** and **EBSCO Host**.

The journal is published online.

Owner: Ali ALEMDAR on Behalf of İstanbul Prof. Dr. Cemil Taşcıoğlu City Hospital

Responsible Manager: İsmail DEMİRKALE

European Archives of Medical Research

Formerly Okmeydanı Medical Journal

CONTENTS

ORIGINAL ARTICLES

104 **Primary Versus Secondary Warm Autoimmune Hemolytic Anemia: Clinical, Laboratory, and Treatment Outcomes in A 20-Year Retrospective Study**
Fatma Pinar Acar, Naciye Demirel, Caner Acar, Pelin Kerter, Mehmet Kucuk; İzmir, Türkiye

112 **Optimizing Medial Release Sequence in Varus Knee Arthroplasty: A Stepwise Approach**
Serhat Gurbuz, Mustafa Bugra Ayaz, Niyazi Igde, Hakan Bahar, Ahmet Keskin, Hakan Gurbuz; Ankara, Türkiye

118 **A Comparative Analysis of Mini-Percutaneous Nephrolithotomy and Retrograde Intrarenal Surgery for the Management of 10–20 Mm Pediatric Kidney Stones**
Recep Burak Degirmentepe, Yasir Muhammed Akca, Deniz Gul, Haci Ibrahim Cimen, Muammer Bozkurt, Tuncay Toprak, Fikret Halis; Ankara, Türkiye

124 **Evaluation of Attitudes, Knowledge Levels, and Perceived Quality of Nutritional Care Among Nurses Working in Internal Medicine and Intensive Care Clinics: A Descriptive Cross-Sectional Study**
Nese Iscan Ayyildiz, Sevgi Demir Cam; Trabzon, Türkiye

133 **Impact of Single-Event Multilevel Surgery on Energy Expenditure and Gait in Children With Cerebral Palsy**
Kadir Cevik, Dilan Deniz Yilmaz, Zeynep Altinkaya, Didem Derici Yildirim, Ugur Dal, Abtullah Milcan; İstanbul, Türkiye

139 **Radiological BI-RADS Scoring and Histopathological Results: Correlation and Pitfalls**
Selçuk Cin, Gül Gizem Pamuk, Muhammed Edip İşleyen, Özgecan Gündoğar, Eda Cingöz, Murat Yüce, Merve Cin, Ali Muhammedoğlu, Ahmet Tan Cimilli; İstanbul, Türkiye

148 **Comparison of HALP and Ranson Scores in Predicting Outcomes of Acute Pancreatitis Admitted to the Emergency Department**
Ibrahim Altundag, Alperen Sahin, Metin Oner, Kaan Yusufoglu, Burcu Genc Yavuz, Sinem Dogruyol, Mustafa Ahmet Afacan, Sahin Colak; İstanbul, Türkiye

156 **Neoadjuvant Pertuzumab in Breast Cancer and Associated Prognostic Factors**
Orcun Can, Seval Ay, Muhammed Mustafa Atci; İstanbul, Türkiye

162 **Comparison of the Quality of English and Turkish Youtube Videos Related to Pes Equinovarus**
Tahsin Olgun Bayraktar, Osman Sariyar, Serdar Aki, Nazım Erkurt, Mehmet Selcuk Saygili, Hakan Gurbuz; İstanbul, Türkiye

169 **Differentiating Physical and Mental Fatigue in Pregnancy: A Chalder-Based Quantitative Analysis**
Elif Ucar, Melih Bestel; İstanbul, Türkiye

174 **Nurses' Attitudes Toward Patients with Chronic Pain**
Özge Demir, Ezgi Demirturk Selcuk; Erzincan, Türkiye

183 **The Importance of Screen Exposure in Cases Diagnosed with Enuresis**
Ozgur Ozdemir Simsek, Gamze Yakin, Hasan Turan, Hasan Akin Tahillioglu; İzmir, Türkiye

European Archives of Medical Research

Formerly Okmeydanı Medical Journal

CONTENTS

ORIGINAL ARTICLES

188 **Machine Learning–Based Ultrasound Radiomics for Differentiating Benign and Malignant Skin Lesions: A Comparative Feature-Set and Model Performance Analysis**
Hakan Ayyildiz; İzmir, Türkiye

197 **Impact of Deep Learning–Based Real-Time Polyp Detection on Adenoma Detection Rate and Concurrent Endoscopist-Performed Optical Biopsy Accuracy: Clinical Validation in Routine Colonoscopy**
Yavuz Ozden, Ferhat Omurca, Dilek Tekis, Oktay Bulur, Nuh Mehmet Buyukberber; Kayseri, Türkiye

206 **Premature Adrenarche Followed to Adolescence: PCOS Risk and Phenotype**
Didem Helvacioglu, Aylin Tugba Canbaz, Busra Gurpinar Tosun, Tulay Guran; Istanbul, Türkiye

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2025.68926
Eur Arch Med Res 2026;42(2):104-111

Primary Versus Secondary Warm Autoimmune Hemolytic Anemia: Clinical, Laboratory, and Treatment Outcomes in A 20-Year Retrospective Study

Fatma Pinar Acar,[1] Naciye Demirel,[2] Caner Acar,[1] Pelin Kerter,[3] Mehmet Kucuk[4]

[1]Department of Medical Oncology, Ege University Hospital, Izmir, Turkiye
[2]Department of Hematology, Prof. Dr. Cemil Tascioglu City Hospital, Istanbul, Turkiye
[3]Department of Internal Medicine, Mardin Training and Research Hospital, Mardin, Turkiye
[4]Department of Nephrology, Prof. Dr. Cemil Tascioglu City Hospital, Istanbul, Turkiye

ABSTRACT

Objective: This study aimed to compare demographic characteristics, laboratory parameters, treatment modalities, and clinical outcomes between adult patients with primary and secondary warm autoimmune hemolytic anemia (wAIHA).

Materials and Methods: This retrospective study included 85 adult patients (57 with primary and 28 with secondary wAIHA) diagnosed between January 2000 and December 2019. Data on baseline hemolytic markers, therapies (corticosteroids, splenectomy, and rituximab), and outcomes were reviewed. The mean age was 53.0±16.2 years, and 74.1% of the patients were female. Patients were assessed for treatment responses and mortality.

Results: No significant differences were observed between the primary and secondary wAIHA groups in terms of baseline hemoglobin, reticulocyte count, lactate dehydrogenase, or bilirubin. However, haptoglobin levels were significantly lower in the primary wAIHA group (p=0.027). Corticosteroid response rates exceeded 90% in both groups, though relapse occurred in 29.4% of responders. Splenectomy led to remission in all secondary and 92% of primary wAIHA cases. Rituximab was administered in only one patient due to reimbursement limitations. Overall mortality was 15.3%, with no significant difference between the primary and secondary groups (p>0.05).

Conclusion: Primary and secondary wAIHA patients show similar clinical and therapeutic profiles. Corticosteroids were highly effective as first-line therapy. Splenectomy showed high response rates, while rituximab usage was limited. Despite clinical differences such as lower haptoglobin in primary wAIHA, mortality rates did not differ significantly.

Keywords: Autoimmune hemolytic anemia, Glucocorticoid, Recurrence, Rituximab, Splenectomy

Cite this article as: Acar FP, Demirel N, Acar C, Kerter P, Kucuk M. Primary Versus Secondary Warm Autoimmune Hemolytic Anemia: Clinical, Laboratory, and Treatment Outcomes in A 20-Year Retrospective Study. Eur Arch Med Res 2026;42(2):104-111.

Address for correspondence: Fatma Pinar Acar. Department of Medical Oncology, Ege University Hospital, Izmir, Türkiye
E-mail: pinar-337@hotmail.com **ORCID ID:** 0009-0000-2749-8732

Submitted: 31.07.2025 **Revised:** 27.10.2025 **Accepted:** 20.12.2025 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

INTRODUCTION

Autoimmune hemolytic anemia (AIHA) arises when the immune system produces antibodies directed against the patient's own erythrocytes, resulting in a rare form of anemia. AIHA is a rare anemia caused by autoantibodies against the patient's own erythrocytes.[1] AIHA is subdivided into three main subtypes according to the thermal activity of the autoantibodies: Warm AIHA (wAIHA), cold AIHA (cold agglutinin disease, cAIHA), and mixed/atypical AIHA. wAIHA is the most common, accounting for approximately 60–70% of all AIHA cases; cold AIHA represents about 15–20%, while mixed and atypical forms are considerably less frequent. Based on thermal activity, AIHA is classified into wAIHA, cold agglutinin disease (cAIHA), and mixed/atypical forms; wAIHA accounts for ~60–70% of cases, cold forms for ~15–20%, and mixed/atypical forms are considerably less frequent and comprise the remainder.[2] In wAIHA, the autoantibodies – predominantly of the immunoglobulin (Ig)G class – are active at 37°C and mediate erythrocyte destruction by macrophage phagocytosis via IgG–Fc receptors in the spleen. Some IgG antibodies may also activate the complement cascade, thereby contributing to hemolysis.[3]

AIHA may be classified as primary or secondary. Primary AIHA is diagnosed when no underlying cause is identified in roughly half of adult patients, whereas secondary AIHA is most often associated with hematological malignancies such as chronic lymphocytic leukemia and non-Hodgkin lymphoma, or with autoimmune disorders such as systemic lupus erythematosus (SLE). Less commonly, immunodeficiencies, infections (e.g., Mycoplasma pneumoniae and human mmunodeficiency virus [HIV]), certain medications (e.g., methyldopa and fludarabine), and solid tumors may give rise to secondary AIHA.[2,4] The diagnosis and monitoring of AIHA rely on laboratory markers of hemolysis. Elevations in reticulocyte count, serum lactate dehydrogenase (LDH), total and indirect bilirubin, together with reduced haptoglobin levels, are the principal indicators of active hemolysis. The direct antiglobulin test (DAT) detects IgG, IgM, and/or complement components bound to the erythrocyte surface. Although a positive DAT in the setting of hemolysis suggests an immune-mediated etiology, a definitive diagnosis of AIHA requires integration of DAT results with clinical findings and other laboratory tests.[5] Etiologically, AIHA is defined as primary when no underlying cause is identified and secondary when associated with another condition – most commonly lymphoproliferative malignancies (e.g., chronic lymphocytic leukemia and non-Hodgkin lymphoma) or systemic autoimmune disease such as SLE; less frequently, immunodeficiencies, infections (e.g., *Mycoplasma pneumoniae* and HIV), medications (e.g., methyldopa and fludarabine), and solid tumors are implicated.[2,4] Diagnosis integrates clinical and laboratory evidence of hemolysis – reticulocytosis; elevated LDH and bilirubin; reduced haptoglobin – together with a positive DAT detecting erythrocyte-bound immunoglobulin and/or complement.[5]

In warm AIHA, corticosteroids remain the standard first-line therapy.[6] Adult patients typically commenced on oral prednisolone at 1–1.5 mg/kg daily; those who respond are usually weaned over 2–3 weeks and the course discontinued within 3–6 months. Lack of a meaningful response by week 3 is considered an indication to discontinue steroids.[4,7,8] Although around 80% of patients initially respond to corticosteroids, more than 50% experience relapse within the 1st year.[9] For patients who are refractory to or dependent on steroids, splenectomy has long been the conventional second-line option in wAIHA, particularly in primary cases, yielding response rates of 60–70% and complete remission in up to 40% of patients. Despite risks of post-operative infection and thrombosis, splenectomy remains a valuable intervention when appropriately selected.[4,10] Since the early 2000s, rituximab – an anti-CD20 monoclonal antibody – has emerged as an alternative to splenectomy in secondline therapy for wAIHA. Studies by Birgens et al.[7] in 2013 and Michel et al.[8] in 2017 demonstrated that adding rituximab to first-line glucocorticoids prolongs remission and improves response rates compared with glucocorticoids alone. Current guidelines recommend rituximab for patients unresponsive to steroids, whilst tailoring the decision for splenectomy based on individual clinical factors. No controlled trials have directly compared rituximab and splenectomy.[4,10] Other immunosuppressants such as azathioprine, ciclosporin, and mycophenolate mofetil may be employed following failure of second-line therapy.[11,12]

Most data on treatment responses and clinical courses in primary versus secondary wAIHA derive from retrospective series with limited patient numbers. In the secondary setting, the heterogeneity of underlying conditions – ranging from hematological malignancies to autoimmune diseases such as SLE – hampers standardization of therapeutic strategies and prediction of clinical outcomes.[4,6,13]

The aim of this study was to compare the demographic, clinical, and laboratory features as well as treatment responses of adult patients diagnosed with warm AIHA of primary versus secondary etiology, thereby elucidating potential differences in management and outcomes.

Adults typically start oral prednisolone 1–1.5 mg/kg/day with early assessment for response and tapering in responders; lack of meaningful improvement by week 3 supports discontinuation.[4,7,8] Initial responses approach 80%, yet >50% relapse within the 1st year.[9] For steroid-refractory or -dependent disease, splenectomy has historically yielded

60–70% responses (complete remission up to 40%) but carries infectious and thrombotic risks.[4,10] Since the early 2000s, rituximab has become a widely used second-line option and, when added to first-line glucocorticoids, can improve response rates and prolong remission;[7,8] other immunosuppressants (azathioprine, cyclosporine, and mycophenolate mofetil) are considered after failure of second-line therapy.[11,12]

Despite these options, comparative evidence for primary versus secondary wAIHA remains limited and largely retrospective, with heterogeneous secondary etiologies complicating standardization and prognostication.[4,6,13] Accordingly, this study compares the demographic, clinical, and laboratory features, as well as treatment responses, of adult patients with primary versus secondary wAIHA to delineate differences in disease behavior and management outcomes.

MATERIALS AND METHODS

This single-center retrospective study was conducted at the Departments of Internal Medicine and Hematology of Prof. Dr. Cemil Taşçıoğlu, City Hospital, affiliated with the University of Health Sciences, between January 2000 and December 2019. Medical records of 85 adult patients diagnosed with warm AIHA were reviewed. The study was performed in accordance with the Declaration of Helsinki and Good Clinical Practice guidelines, with approval from the Clinical Research Ethics Committee of Prof. Dr. Cemil Taşçıoğlu City Hospital (approval no: 2020/212; decision date: June 02, 2020; document no: 48670771514.10).

Demographic and clinical data recorded included age, sex, primary or secondary AIHA classification, concomitant diseases, baseline hemoglobin (Hb), reticulocyte percentage, LDH, total and indirect bilirubin, alanine and aspartate transaminases, haptoglobin, DAT results, and anticardiolipin IgM status. Data on specific antibody isotypes and complement binding were unavailable, as only a polyspecific anti-human globulin reagent is used in our institution's blood bank.

First-line therapy consisted of methylprednisolone or prednisolone. Patients who failed to respond or relapsed proceeded to second-line treatment, which comprised splenectomy or rituximab. Two patients were lost to follow-up before the first-line response could be assessed; hence, response analysis was performed on 83 patients. Complete response was defined as Hb ≥12 g/dL with resolution of hemolytic markers (normalization of bilirubin, LDH, and haptoglobin). Partial response was defined as Hb ≥10 g/dL with at least a 2 g/dL increase over baseline, despite persistence of mild hemolysis. Refractory disease was defined as failure to achieve the expected Hb rise by week 2 and absence of significant improvement by week 3. Relapse was defined as a drop in Hb and recurrence of hemolytic markers after initial complete or partial response.[5]

Statistical Analysis

Statistical analyses were performed using IBM Statistical Package for the Social Sciences (SPSS) Statistics 22 (IBM SPSS, Türkiye). Normality was tested by the Shapiro–Wilk test. Continuous variables are presented as mean±standard deviation. Comparisons between two groups used Student's *t*-test for normally distributed parameters and the Mann–Whitney U test otherwise. Categorical variables are expressed as numbers and percentages and compared using Fisher's Exact test, the Fisher–Freeman–Halton test, or the Chi-square test with Yates' correction as appropriate. $P<0.05$ was considered statistically significant.

RESULTS

Demographic and Clinical Characteristics of Patients

A total of 85 patients met the inclusion criteria. Ages ranged from 18 to 87 years, with a mean of 53.0±16.2 years. 65 (74.1%) patients were female, with no appreciable difference in sex distribution between the primary and secondary AIHA groups. Of the cohort, 57 patients (67.1%) had primary AIHA and 28 (32.9%) had secondary AIHA. In the secondary group, underlying hematological malignancies were most common (46.4%), followed by autoimmune diseases (28.6%), solid tumors (14.3%), drug-related AIHA (7.1%), and infection-related AIHA (3.6%) (Table 1).

Table 1. Underlying conditions associated with secondary AIHA.

	n	%
Hematologic malignancy	13	46.4
Chronic lymphocytic leukemia	6	21.4
Hodgkin lymphoma	3	10.7
Non-Hodgkin lymphoma	2	7.1
Multiple myeloma	1	3.6
Acute myeloid leukemia	1	3.6
Solid tumors	4	14.3
Autoimmune diseases	8	28.6
Systemic lupus erythematosus	5	17.9
Rheumatoid arthritis	2	7.1
Primary biliary cholangitis	1	3.6
Other causes	3	10.7
Drugs	2	7.1
Infection	1	3.6

AIHA: Autoimmune hemolytic anemia

Comorbidities were present in 45 patients (52.9%), most frequently hypertension (27%), diabetes mellitus (9.4%), and chronic kidney disease (10.6%). 85% of patients exhibited at least one symptom at presentation, most commonly fatigue (64.7%), palpitations (12.9%), and dizziness (17.6%) (Table 2). No significant differences in demographics or symptom frequency were observed between primary and secondary groups (p>0.05).

Laboratory comparisons revealed that haptoglobin levels were significantly lower in primary AIHA (90.9%) than in secondary AIHA (54.5%) (p=0.027). Other hemolytic markers did not differ significantly between groups (p>0.05) (Table 3).

Treatment and response rates

First-line corticosteroid therapy was administered to 85 patients (97.6% received methylprednisolone; 2.4% prednisolone) for a duration of 1–58 months (mean 6.46±6.8 months). However, treatment response could be evaluated in only 83 patients because two were lost to follow-up before response assessment. Seventy-eight patients (94%) responded; 58 (69.9%) with complete and 20 (24.1%) with partial responses; five patients (6%) were refractory. Response rates did not differ significantly between primary (52/56; 92.8%) and secondary (26/27; 96.2%) AIHA (p>0.05). Treatment responses to corticosteroids and splenectomy are summarized in Table 4. Relapse occurred in 23 of 78 responders (29.4%); relapse rates were similar in primary (28.8%) and secondary (30.8%) groups. Of relapsed patients, 12 restarted steroids, and 11 underwent splenectomy. Among refractory cases, three had splenectomy, and one received rituximab; one secondary AIHA patient with concomitant acute myeloid leukemia died before second-line therapy.

Splenectomy was performed in 14 patients (16.4%) as second-line therapy – 12 primary (21%) and two secondary (7.1%) AIHA cases (P=0.027). Eleven primary patients achieved complete and one partial remission; both secondary patients attained complete remission. No refractory cases were noted in either subgroup (p>0.05). Rituximab was used in only one primary AIHA patient, who did not respond; reinitiation of corticosteroids achieved complete remission.

Mortality

During follow-up, 13 patients (15.3%) died; Seven with primary AIHA (12.3%) and six with secondary AIHA (21.4%), with no significant difference in mortality between groups (p>0.05).

DISCUSSION

In this retrospective single-center series of 85 adults with warm AIHA, 67% had primary disease and 33% secondary. Whereas most series report an even distribution of primary

Table 2. Demographic and clinical characteristics of primary and secondary wAIHA patients.

	Primary OIHA (n=57) (%)	Secondary OIHA (n=28) (%)	p
Demographic characteristics			
Age (mean±SD)	52.59±15.91	53.93±16.98	[1]0.723
Gender *n* (%)			
Male	13 (22.8)	9 (32.1)	[2]0.509
Female	44 (77.2)	19 (67.9)	
Comorbidity			
Yes	27 (47.4)	18 (64.3)	[1]0.216
No	30 (52.6)	10 (35.7)	
Hypertension			
Yes	16 (28.1)	7 (25)	[1]0.968
No	41 (71.9)	21 (75)	
Chronic kidney disease			
Yes	7 (12.3)	2 (7.1)	[2]0.711
No	50 (87.7)	26 (92.9)	
Diabetes mellitus			
Yes	6 (10.5)	2 (7.1)	[2]1.000
No	51 (89.5)	26 (92.9)	
Hypothyroidism			
Yes	3 (5.3)	4 (14.3)	[2]0.211
No	54 (94.7)	24 (85.7)	
Presenting symptoms			
Yes	41 (71.9)	17 (60.7)	[1]0.426
No	16 (28.1)	11 (39.3)	
Fatigue			
Yes	39 (68.4)	16 (57.1)	[1]0.435
No	18 (31.6)	12 (42.9)	
Palpitation			
Yes	7 (12.3)	4 (14.3)	[2]1.000
No	50 (87.7)	24 (85.7)	
Dizziness			
Yes	13 (22.8)	2 (7.1)	[2]0.128
No	44 (77.2)	26 (92.9)	
Dyspnea			
Yes	12 (21.1)	2 (7.1)	[2]0.129
No	45 (78.9)	26 (92.9)	

Table 2. Continue.

	Primary OIHA (*n*=57) (%)	Secondary OIHA (*n*=28) (%)	*p*
Jaundice			
Yes	9 (15.8)	5 (17.9)	[2]1.000
No	48 (84.2)	23 (82.1)	
Angina			
Yes	3 (5.3)	0 (0)	[2]0.548
No	54 (94.7)	28 (100)	

Data are presented as *n* (%) for categorical variables and as mean±SD for continuous variables. SD: Standard deviation, wAIHA: Warm autoimmune hemolytic anemia

and secondary AIHA, the predominance of primary cases in our cohort likely reflects underrecognition of secondary etiologies over the 20-year study period.[4,14] Consistent with previous studies, hematological malignancies (46.4%) and autoimmune diseases (28.6%) were the leading secondary etiologies.[15,16] No significant age difference was observed between subgroups, though larger cohort studies have reported older age at onset in secondary AIHA.[17] Female predominance in both primary and secondary groups aligns with existing literature.[18,19]

In our series, the mean Hb level at diagnosis was lower in the secondary AIHA group, reflecting a tendency toward more pronounced anemia at presentation. A similar study reported mean Hb concentrations of 6.3 g/dL in secondary warm AIHA patients versus 7.1 g/dL in primary cases.[16] Although we observed a numerical difference between the two groups, it did not reach statistical significance. No significant differences were identified between primary and secondary AIHA in Hb and reticulocyte counts.[16,20] In our study, haptoglobin levels were the only parameter that differed between the two groups; low haptoglobin was significantly more frequent in the primary warm AIHA cohort. No previous studies have reported such a discrepancy. Although this finding is noteworthy, its interpretation must be cautious given the small sample size and the absence of a clear mechanistic rationale.[15,16]

Corticosteroids remain the cornerstone of first-line therapy in warm AIHA, and in our study, more than 90% of patients in both cohorts responded to prednisolone. Overall, initial response rates were high, with no statistically significant difference between primary and secondary AIHA. Our findings align with numerous reports indicating that 80–90% of wAIHA cases respond to firstline steroids.[17] For example, Rattarittamrong et al.[21] observed a 96% response rate – regardless of whether AIHA was primary or SLEassociated secondary – whereas

Table 3. Laboratory findings of patients with primary and secondary wAIHA.

	Primer OIHA (*n*=57)	Sekonder OIHA (*n*=28)	*p*
Hemoglobin (g/dL)	7.35±2.02	6.63±1.91	0.121
Hematocrit (%)	21.96±6.29	20.45±5.58	0.306
Reticuloscyte percentage	10.72±7.44	9.72±9.28	0.268
LDH (U/L).	679.93±754.69	600.4±561.45	0.720
Total bilirubin (mg/dL)	2.94±2.79	2.45±2.27	0.440
İndirect bilirubin (mg/dL)	2.47±2.47	2.7±1.85	0.281
ALT (U/L).	60.8±239.55	24.27±22.66	0.812
000.812AST (U/L).	44.88±78.18	36±41.63	0.410
Haptoglobin (%)			0.027*
Low	20 (90.9)	6 (54.5)	
High	2 (9.1)	5 (45.5)	
Direct coombs test (%)			0.256
Positive	48 (84.2)	27 (96.4)	
Negative	9 (15.8)	1 (3.6)	

ALT: Alanine aminotransferase; AST: Aspartate aminotransferase; DAT: Direct antiglobulin (Coombs) test; Hb: Hemoglobin; Hct: Hematocrit; LDH: Lactate dehydrogenase; wAIHA: Warm autoimmune hemolytic anemia.
Data are presented as *n* (%) for categorical variables and as mean±standard deviation for continuous variables. Haptoglobin levels were measured in 22 patients with primary AIHA and 11 patients with secondary AIHA, with values <30 mg/dL considered low. Direct and indirect Coombs tests were reported as positive or negative. Student's *t*-test or Mann–Whitney U test was used for continuous variables, and Fisher's exact test was applied for categorical variables.

Table 4. Treatment responses to first- and second-line therapy in primary and secondary wAIHA patients.

Treatment	Response	Primary wAIHA (*n*=56) (%)	Secondary wAIHA (*n*=27)	Total wAIHA (*n*=83)	*p*
Corticosteroid	Complete response	40 (71.4)	18 (66.6)	58 (69.9)	
	Partial response	12 (21.4)	8 (29.6)	20 (24.1)	0.62
	Refractory to treatment	4 (7.2)	1 (3.8)	5 (6.0)	
Splenectomy	Complete response	11 (91.7)	2 (100)	13 (92.9)	1.00
	Partial response	1 (8.3)	0	1 (7.1)	
	Refractory to treatment	0	0	0	

wAIHA: Warm autoimmune hemolytic anemia

Sudulagunta et al.[16] reported response rates of approximately 97% in primary wAIHA and 87% in secondary wAIHA. In our cohort, the vast majority of both primary and secondary cases responded to steroids, and refractory rates were low. This concordance demonstrates that secondary AIHA patients can achieve outcomes comparable to those with primary AIHA, supporting guideline recommendations to employ the same initial treatment strategy in both subtypes.[4,22]

Among patients who initially responded to steroids, relapse occurred in roughly 30–28.8% in the primary group and 30.8% in the secondary group. In the series by Rattarittamrong et al.[21] involving 101 warm AIHA patients, the overall relapse rate after steroids was 49.5% (50% in primary; 48.7% in secondary).[21] It is generally understood that while about 80% of patients respond to steroids initially, only 30–40% maintain durable remission at 1 year.[9] Our observed relapse rates were lower, likely owing to attrition in long-term follow-up over the 20-year period, which may have led to underestimation of true relapse frequency.

In our study, splenectomy was performed as second-line therapy in 14 patients, 12 of whom had primary AIHA and 2 of whom had secondary AIHA. The rate of splenectomy was significantly higher in the primary AIHA group (21%) compared with the secondary group (7.1%). Among those who underwent splenectomy, 11 primary AIHA patients achieved complete remission, and 1 achieved a partial response, whereas both secondary AIHA patients attained complete remission. Although response rates were high in both cohorts, splenectomy was significantly favored for primary AIHA patients in our series. International consensus reports indicate that response to splenectomy in secondary AIHA may be comparable to that in primary cases; however, they emphasize that splenectomy should be avoided where possible in secondary AIHA because of the increased risk of infectious and thrombotic complications associated with comorbid conditions.[4] Conversely, splenectomy remains a well-established secondline option in primary AIHA, with literature reporting 60–70% response rates. Our results are consistent with these data, and the wider adoption of laparoscopic techniques has further reduced operative risks, underscoring splenectomy's continued role in appropriately selected patients.[11,23,24]

Rituximab has been used since the early 2000s as a second-line therapy for steroid-refractory or -dependent wAIHA and, in some studies, in combination with steroids as first-line treatment.[7,8,10] However, in our series, rituximab was administered to only one primary AIHA patient, who did not respond. This low utilization likely reflects the drug's recent introduction into practice during the study period and restrictive reimbursement policies that limited access.

In our cohort, mortality occurred in 13 of the 85 patients: Seven in the primary AIHA group and six in the secondary AIHA group. This difference did not reach statistical significance. Although some series report higher mortality in secondary AIHA, others have found no significant difference between primary and secondary cases. Our data demonstrate that the numerically greater death rate in the secondary subgroup was not statistically significant.[16,19]

Limitations include the retrospective design, potential incomplete evaluation of secondary causes, variable follow-up durations, and missing data for certain laboratory parameters. Despite these constraints, our findings support comparable treatment responses in primary and secondary warm AIHA and underscore the need for standardized evaluation and management protocols tailored to underlying etiologies.

CONCLUSION

In this retrospective study of 85 adults with wAIHA, primary and secondary subgroups exhibited comparable demographic characteristics, laboratory profiles, and treatment responses. First-line corticosteroid therapy produced high initial

response rates (>90%) in both cohorts, with relapse occurring in approximately 30%. Splenectomy demonstrated robust efficacy as a secondline intervention, while rituximab use was minimal due to reimbursement constraints. Mortality rates did not differ significantly between subgroups, despite a numerically higher rate in secondary cases. These findings suggest that primary and secondary wAIHA follow similar clinical courses and respond similarly to current treatment algorithms.

DECLARATIONS

Ethics Committee Approval: The study was approved by Prof. Dr. Cemil Taşçıoğlu City Hospital Ethics Committee (No: 48670771 514.10, Date: 02/06/2020).

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: No artificial intelligence tools were used in the preparation of this manuscript.

Authorship Contributions: Concept – ND, MK; Design – FPA, ND; Supervision – FPA, ND; Fundings – MK; Materials – FPA; Data collection &/or processing – PK, CA; Analysis and/or interpretation – CA; Literature search – CA; Writing – FPA, PK.

Acknowledgment: This study is based on the internal medicine specialty thesis of Dr. Fatma Pınar Açar.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Zanella A, Barcellini W. Treatment of autoimmune hemolytic anemias. Haematologica 2014;99:1547–54.
2. Loriamini M, Cserti-Gazdewich C, Branch DR. Autoimmune hemolytic anemias: classifications, pathophysiology, diagnoses and management. Int J Mol Sci 2024;25:4296.
3. Mulder FVM, Evers D, de Haas M, Cruijsen MJ, Bernelot Moens SJ, Barcellini W, et al. Severe autoimmune hemolytic anemia; epidemiology, clinical management, outcomes and knowledge gaps. Front Immunol 2023;14:1228142.
4. Jäger U, Barcellini W, Broome CM, Gertz MA, Hill A, Hill QA, et al. Diagnosis and treatment of autoimmune hemolytic anemia in adults: Recommendations from the First International Consensus Meeting. Blood Rev 2020;41:100648.
5. Hill QA, Stamps R, Massey E, Grainger JD, Provan D, Hill A, et al. The diagnosis and management of primary autoimmune haemolytic anaemia. Br J Haematol 2017;176:395–411.
6. Barcellini W, Fattizzo B. How I treat warm autoimmune hemolytic anemia. Blood 2021;137:1283–94. Erratum in: Blood 2023;141:438–9.
7. Birgens H, Frederiksen H, Hasselbalch HC, Rasmussen IH, Nielsen OJ, Kjeldsen L, et al. A phase III randomized trial comparing glucocorticoid monotherapy versus glucocorticoid and rituximab in patients with autoimmune haemolytic anaemia. Br J Haematol 2013;163:393–9.
8. Michel M, Terriou L, Roudot-Thoraval F, Hamidou M, Ebbo M, Le Guenno G, et al. A randomized and double-blind controlled trial evaluating the safety and efficacy of rituximab for warm auto-immune hemolytic anemia in adults (the RAIHA study). Am J Hematol 2017;92:23–7.
9. Berentsen S, Barcellini W. Autoimmune hemolytic anemias. N Engl J Med 2021;385:1407–19.
10. Reynaud Q, Durieu I, Dutertre M, Ledochowski S, Durupt S, Michallet AS, et al. Efficacy and safety of rituximab in autoimmune hemolytic anemia: A meta-analysis of 21 studies. Autoimmun Rev 2015;14:304–13.
11. Hill A, Hill QA. Autoimmune hemolytic anemia. Hematology Am Soc Hematol Educ Program 2018;2018:382–9.
12. Cavallaro F, Barcellini W, Fattizzo B. Antibody based therapeutics for autoimmune hemolytic anemia. Expert Opin Biol Ther 2023;23:1227–37.
13. Lechner K, Jäger U. How I treat autoimmune hemolytic anemias in adults. Blood 2010;116:1831–8.
14. Alonso HC, Manuel AV, Amir CG, Sergio RR, Allan P, Xavier LK, et al. Warm autoimmune hemolytic anemia: experience from a single referral center in Mexico City. Blood Res 2017;52:44–9.
15. Roumier M, Loustau V, Guillaud C, Languille L, Mahevas M, Khellaf M, et al. Characteristics and outcome of warm autoimmune hemolytic anemia in adults: New insights based on a single-center experience with 60 patients. Am J Hematol 2014;89:E150–5.
16. Sudulagunta SR, Kumbhat M, Sodalagunta MB, Settikere Nataraju A, Bangalore Raja SK, Thejaswi KC, et al. Warm autoimmune hemolytic anemia: clinical profile and management. J Hematol 2017;6:12–20.
17. Cherif H, Cai Q, Crivera C, Leon A, Rahman I, Leval A, et al. Overall survival and treatment patterns among patients with warm autoimmune hemolytic anemia in sweden: a nationwide population-based study. Eur J Haematol 2025;114:129–38.
18. Tranekær S, Hansen DL, Frederiksen H. Epidemiology of secondary warm autoimmune haemolytic anaemia-a systematic review and meta-analysis. J Clin Med

2021;10:1244.

19. Hansen DL, Möller S, Frederiksen H. Survival in autoimmune hemolytic anemia remains poor, results from a nationwide cohort with 37 years of follow-up. Eur J Haematol 2022;109:10–20.
20. Mueller L, Abunafeesa H, Cabrera Fernandez D, Kuriakose P. Primary versus secondary autoimmune hemolytic anemia: a retrospective analysis of diagnostic patterns and treatment outcomes. Blood 2018;132(Suppl 1):4881.
21. Rattarittamrong E, Eiamprapai P, Tantiworawit A, Rattanathammethee T, Hantrakool S, Chai-Adisaksopha C, et al. Clinical characteristics and long-term outcomes of warm-type autoimmune hemolytic anemia. Hematology 2016;21:368–74.
22. Berentsen S, Fattizzo B, Barcellini W. The choice of new treatments in autoimmune hemolytic anemia: how to pick from the basket? Front Immunol 2023;14:1180509.
23. Cvetković Z, Pantić N, Cvetković M, Virijević M, Sabljić N, Marinković G, et al. The role of the spleen and the place of splenectomy in autoimmune hemolytic anemia-a review of current knowledge. Diagnostics (Basel) 2023;13:2891.
24. Murakhovskaya I. Rituximab use in warm and cold autoimmune hemolytic anemia. J Clin Med 2020;9:4034.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2025.24582
Eur Arch Med Res 2026;42(2):112-117

Optimizing Medial Release Sequence in Varus Knee Arthroplasty: A Stepwise Approach

Serhat Gurbuz,[1] Mustafa Bugra Ayaz,[1] Niyazi Igde,[1] Hakan Bahar,[1] Ahmet Keskin,[1] Hakan Gurbuz[2]

[1]Department of Orthopedics and Traumatology, University of Health Sciences, Metin Sabanci Baltalimani Bone Diseases Training and Research Center, Istanbul, Turkiye
[2]Department of Orthopedics and Traumatology, University of Health Sciences, Prof. Dr. Cemil Tascioglu City Hospital, Istanbul, Turkiye

ABSTRACT

Objective: Achieving optimal ligament balance in total knee arthroplasty (TKA) requires a precise sequence of medial release techniques. While medial collateral ligament (MCL) release is often considered the first step, there is no consensus regarding the subsequent steps and their effectiveness. This study examines a stepwise medial release protocol for varus knees to determine its efficacy in improving intraoperative ligament balance and post-operative outcomes.

Materials and Methods: A retrospective analysis was conducted on 60 patients (39 females, 21 males) who underwent primary TKA for varus deformity between January 2020 and December 2022. A structured medial release sequence was applied, consisting of the superficial MCL, deep MCL, posterior capsule, Pes Anserinus, and increased tibial resection. Gap balancer measurements were recorded at 0° and 90° flexion. Statistical analyses included paired t-tests and Chi-square tests to assess the impact of each release step on ligament balance.

Results: At 0° flexion, 85% of patients achieved optimal medial ligament balance after completing the release sequence. At 90° flexion, the optimal balance was achieved in 80% of cases. The mean Knee Society Score improved from 45.2±10.5 preoperatively to 88.7±8.9 at the 12-month follow-up ($p<0.05$). Range of motion increased from 95°±15° to 120°±10° ($p<0.05$).

Conclusion: The stepwise release protocol effectively improved medial ligament balance without excessive soft-tissue release, reducing the risk of post-operative instability. The sequence provided a reproducible and systematic approach, facilitating consistent intraoperative decision-making. The proposed stepwise medial release protocol effectively balances ligaments in varus knees during TKA, resulting in improved clinical outcomes and reduced post-operative complications. Future studies should assess the protocol's applicability in diverse patient populations and compare its effectiveness with robotic-assisted techniques.

Keywords: Arthroplasty, Knee, Knee prosthesis, Medial release, Total knee arthroplasty

Cite this article as: Gurbuz S, Ayaz MB, Igde N, Bahar H, Keskin A, Gurbuz H. Optimizing Medial Release Sequence in Varus Knee Arthroplasty: A Stepwise Approach. Eur Arch Med Res 2026;42(2):112-117.

Address for correspondence: Serhat Gurbuz. Department of Orthopedics and Traumatology, University of Health Sciences, Metin Sabancı Baltalimanı Bone Diseases Training and Research Center, İstanbul, Türkiye

E-mail: gurbuzserhat@outlook.com **ORCID ID:** 0000-0002-6120-3325

Submitted: 12.05.2025 **Revised:** 01.07.2025 **Accepted:** 21.07.2025 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

INTRODUCTION

Gonarthrosis is one of the most common degenerative joint diseases, particularly in the elderly population, and is characterized by pain, restricted movement, and functional loss in the knee joint. Total knee arthroplasty (TKA) is widely employed as an effective surgical intervention in advanced-stage knee osteoarthritis to alleviate pain and improve quality of life. However, knee deformities, particularly varus deformity, pose significant challenges in surgical planning and execution.

Varus knee deformity leads to degeneration in the medial compartment due to increased loading on the medial structures and is often accompanied by medial soft-tissue contractures.[1] During TKA, the release of these contractures (medial release) is crucial for achieving proper mechanical alignment and maintaining prosthetic stability. However, there are varying approaches in the literature regarding the sequence, targeted structures, and extent of medial release.

Soft-tissue balancing procedures performed during surgery play a fundamental role in achieving neutral alignment of the knee.[2] These procedures are typically carried out in a stepwise manner, targeting structures such as the deep medial collateral ligament (dMCL), osteophyte removal, Pes Anserinus insertion site, semimembranosus tendon, posteromedial capsule, and superficial medial collateral ligament (sMCL).[3] It is crucial to accurately determine the release sequence of each structure, as excessive release may increase the risk of medial instability, while inadequate release may result in difficulties in prosthesis placement.[4]

Recent advancements in surgical techniques and the increasing use of computer-assisted systems have enabled a more systematic approach to soft-tissue balancing strategies. However, the surgeon's experience and patient-specific anatomical variations remain critical in achieving optimal outcomes. Therefore, establishing an evidence-based, structured, and reproducible release sequence is essential for clinical success.

In this study, the stepwise approach to medial release in patients undergoing primary TKA for varus knees will be thoroughly examined. Various approaches and techniques reported in the literature will be compared to propose the most effective and safe sequence. The aim is to contribute to intraoperative decision-making processes for surgeons while also improving post-operative functional outcomes.

MATERIALS AND METHODS

Study Design and Patient Selection

A retrospective analysis was conducted on 60 consecutive patients with varus deformity who underwent primary TKA between January 2020 and December 2022. Inclusion criteria included primary osteoarthritis with varus deformity >10° and intact collateral ligaments. Exclusion criteria included previous knee surgery, severe valgus deformity, inflammatory arthritis, and varus deformity over 30°. This study was conducted in accordance with the principles outlined in the Declaration of Helsinki, ensuring the protection of participants' rights and welfare throughout the research process. Ethical approval for this study was obtained from the local institutional review board (Decision-22/146), and all procedures were carried out following the approved protocol.

Surgical Technique

All procedures were performed by a single surgeon using a medial parapatellar approach. After bone cuts, ligament balance was assessed using a gap balancer at 0° and 90° of flexion. Medial tightness was scored as follows:

- 1 point: Lateral entry only.
- 2 points: Medial and lateral entry with medial tightness.
- 3 points: Optimal medial laxity.

A stepwise medial release sequence was applied based on gap measurements:

1. MCL release at sagittal midline
2. Removal of medial tibial osteophytes
3. Posterior capsular release and posterior osteophyte removal
4. Pes anserinus tenotomy
5. Increased tibial resection

Data Collection

Intraoperative gap measurements were recorded at both 0° and 90° of flexion for each step of the release sequence. The data were categorized by gender (male: "m", female: "f") and the effectiveness of each release step was scored as described above.

Statistical Analysis

Data were analyzed using the Statistical Package for the Social Sciences version 25.0. Continuous variables, such as pre-operative and post-operative gap balancer scores and range of motion (ROM), were compared using paired *t*-tests to assess the effectiveness of the medial release steps in both 0° and 90° flexion. The effectiveness of each release step in achieving optimal medial ligament balance was further evaluated through chi-square tests to compare the distribution of gap

balancer scores across different patient groups and release steps. In addition, the overall improvement in Knee Society Scores (KSS) from pre-operative to post-operative follow-up was assessed using paired *t*-tests, with statistical significance set at $p<0.05$. Gender-based differences in achieving optimal ligament balance were also analyzed using Chi-square tests to determine potential disparities between male and female patients.

RESULTS

Demographics

The study included 60 patients (39 females, 21 males) with a mean age of 68.5±7.2 years. The mean pre-operative varus deformity was 12.3°±3.1° (Table 1).

Intraoperative Findings at 0° Flexion

The stepwise release sequence significantly improved medial ligament balance at 0° flexion ($p<0.05$) (Table 2). The distribution of gap balancer scores at each step was as follows:

- After dMCL release: 45% achieved score 3
- After osteophyte removal: 65% achieved score 3
- After posterior release: 75% achieved score 3
- After pes anserinus tenotomy: 80% achieved score 3
- After ıncreased tibial resection: 85% achieved score 3.

Intraoperative Findings at 90° Flexion

At 90° flexion, the stepwise release sequence also showed significant improvement in medial ligament balance ($p<0.05$). The distribution of gap balancer scores at each step was as follows:

- After dMCL release: 40% achieved score 3
- After osteophyte removal: 60% achieved score 3
- After posterior release: 70% achieved score 3
- After pes anserinus tenotomy: 75% achieved score 3
- After ıncreased tibial resection: 80% achieved score 3.

Gender-based Analysis

- Female patients (f): 85% achieved optimal balance (score 3) at 0° flexion, and 80% at 90° flexion
- Male patients (m): 80% achieved optimal balance (score 3) at 0° flexion, and 75% at 90° flexion.

Post-operative Outcomes

At 12-month follow-up, the mean KSS improved from 45.2±10.5 preoperatively to 88.7±8.9 postoperatively ($p<0.05$). The mean ROM improved from 95°±15° to 120°±10° ($p<0.05$) (Table 3).

DISCUSSION

The objective of this study is to evaluate the adequacy of medial release using a simple intraoperative method and to propose a protocol regarding the sequence of medial structure release, which remains a debated topic. Following the completion of bone resections and osteophyte excisions, we consider the order of medial structures to be released as the dMCL, posterior-oblique ligament (POL), posteromedial capsule, sMCL, Pes Anserinus, and Semimembranosus muscles. In addition, as a secondary outcome, we believe that the gap balancer utilized during surgery is sufficient for assessing ligament balance.

In arthritic varus knees, the tension of the medial soft tissues and the laxity of the lateral soft tissues are directly related to the degree of varus deformity. The objective of TKA is to achieve a rectangular gap that is equally balanced in both flexion and extension. Thus, releasing tight structures appears to be a necessary intervention. However, excessive release can result in severe complications, such as knee instability. Furthermore, recent advancements in robotic-assisted surgery have provided a more comfortable postoperative course compared to conventional techniques, particularly during

Table 1. Age distribution and outcomes analysis.

Age group	Gender	Number of patients	Mean KSS score (Postop)	Mean ROM (Degrees)
<60	Male	12	85	115
<60	Female	15	87	118
60–70	Male	10	90	120
60–70	Female	18	89	122
>70	Male	9	83	112
>70	Female	12	86	115

KSS: Knee society scores; ROM: Range of motion.

Table 2. Effect of release steps on medial balance.

Release step	Patients achieving optimal balance (%)	Average gap balancer score (0°)	Average gap balancer score (90°)
Medial collateral ligament release	45	2.1	2.0
Osteophyte removal	65	2.5	2.3
Posterior capsule release	75	2.8	2.6
Pes anserinus tenotomy	80	3.0	2.8
Increased tibial resection	85	3.0	3.0

Table 3. ***Complications based on release steps***

Number of release steps	Complications (%)	Instability cases	Joint stiffness cases
1–2 Steps	10	2	1
3–4 Steps	15	3	2
All 5 Steps	20	5	3

the initial 6-month period. This postoperative comfort has been attributed to the limited intervention on ligamentous structures. Therefore, achieving adequate but minimal release may lead to more favorable outcomes.

Furthermore, it should be noted that during TKA in varus knees, the dMCL is often inadvertently released as part of the standard surgical exposure. Specifically, when the medial meniscus is excised, the dMCL, which is anatomically connected to the joint capsule and the medial meniscus, is also necessarily disrupted. Therefore, this step typically precedes the removal of medial tibial osteophytes. In this context, the beginning sequence described in our study reflects the practical surgical order rather than a deliberate soft-tissue balancing strategy. Nonetheless, we agree that osteophyte removal plays a significant role in medial tension release and have clarified this point accordingly.

First and foremost, confirming medial-lateral ligament balance during surgery itself presents a significant contradiction. There are numerous perspectives on this matter, and several instruments developed by relevant companies are available, including the Zimmer offset-repo, Stryker's devices, and those from Smith and Nephew, all of which have been extensively studied and reported in the literature.[5,6] In addition, there is a study that assesses ligament balance intraoperatively using a computer-assisted system.[7]

Therefore, in the first stage of our study, we developed a simple, cost-effective, and easily applicable method specific to our approach. In this method, we utilized a carbon gap balancer, which is commonly available in every TKA set, to define a three-stage laxity scoring system. We assigned scores of 1, 2, and 3 based on the procedures performed during surgery, both in extension and at 90° flexion, after completing the bone cuts. Given the accessibility and cost-effectiveness of this method, we believe it has the potential for widespread clinical application.

Most orthopedic surgeons prefer to use the thinnest possible insert, as a smaller insert is associated with reduced wear.[4] We consider the size of the gap balancer to be irrelevant, as the primary objective in mechanically aligned TKA is not only to release the tight medial structures but also to balance the lax lateral structures, ultimately achieving medial-lateral ligament balance. Therefore, the insert size is not a determining factor for us. Even if a larger size is tested, we believe that the clinical interpretation should consider both medial and lateral engagement, which is ultimately influenced by the surgeon's intraoperative experience.

In their study, Keggi et al. [8] reported that the use of a digital balancebot device during robotic knee arthroplasty improved medial-lateral ligament balance at the end of the procedure. Although robotic knee arthroplasty ensures more precise bone cuts, progress in ligament balance remains limited. [9] Regarding ligament balance, kinematic alignment in TKA does not recommend medial release, preserving the patient's native soft-tissue balance.[10] Kinematically aligned TKA using the measured resection technique requires a high level of surgical expertise and presents significant challenges in non-robotic procedures. In our study, as we performed non-robotic surgeries, we did not opt for kinematic alignment, and therefore, ligament balance was achieved manually.[11] Supporting this notion, the study by Song demonstrated that sensor-assisted TKA is not superior to manual techniques.

[12] A study utilizing intraoperative load distribution sensors demonstrated that the sensor was found to be ineffective.[13]

In a study conducted by Geller et al., [14] it was reported that collateral ligament balancing did not differ postoperatively or at the 1-year follow-up following knee arthroplasty. We disagree with this perspective, as the post-operative follow-up period in this study was relatively short. In our clinical practice, we have observed a strong correlation between the intraoperative ligamentous balance achieved during surgery and the postoperative patient-reported outcome scores in patients we have followed over a longer duration. Another study suggests that robotic-assisted bone cuts performed without intervening in the soft tissues achieve more accurate intraoperative ligament balance.[15]

Recent advancements suggest that surgeons should adopt a conservative approach to ligament releases, as each ligament is proprioceptively effective.[16] Both the surgeon and the patient may require these ligaments in the future. In this study, our objective was to assess the optimal starting point for ligament releases and to determine which specific ligaments should be released to achieve perioperative balance in the most conservative manner.

A cadaveric study demonstrated that medial release, particularly of the sMCL, significantly affects anterior stability in ligament-preserving TKA. The study also emphasizes that surgeons should avoid excessive medial release and adopt a minimalist approach.[17] Recent trends indicate that minimal soft-tissue release has yielded satisfactory mid- to long-term outcomes.[18]

There are numerous proposed methods for medial release, each with its own recommended sequence. In our practice, we perform the classic dMCL release, sMCL release up to the Pes Anserinus, medial tibial osteophyte removal, posterior capsulotomy, and extension of the tibial cut. We do not utilize other methods, such as needle-assisted MCL release, which have been described in the literature, as we consider them to be ineffective.[19]

The medial structures to be released can be categorized as both static and dynamic. The static structures include the sMCL and dMCL, POL, and the posteromedial capsule. The dynamic structures encompass the Pes Anserinus and Semimembranosus muscles. In severe varus deformities, all of these structures may need to be released. However, in cases with moderate varus deformities, a stepwise and sequential approach is the most reasonable strategy. This approach ensures post-operative comfort and prevents instability issues.

Our findings have important clinical implications for surgeons performing TKA in varus-deformed knees. The stepwise medial release sequence based on intraoperative gap measurements provides a systematic and reproducible method for achieving optimal ligament balance. This approach can help reduce post-operative complications, such as instability and accelerated polyethylene wear, and improve patient outcomes.

In our study, we assessed medial-lateral ligament balance using a gap gauge both in extension and flexion. Our objective was to determine when, in which patients, and in what sequence these releases should be performed. In addition, we aimed to explore whether this method could potentially reduce mid-flexion instability, which remains one of the major challenges in knee arthroplasty.

Limitations

The limitations of our study include its retrospective design and the absence of robotic navigation methods. Second, the study was conducted at a single center, which may limit the generalizability of the findings. While we acknowledge that robotic navigation may provide more precise and accurate ligament balancing, we believe that its routine use is challenging due to its high cost and limited accessibility.

CONCLUSION

In varus knees, we followed a general approach to medial release, starting with osteotomy. Following the release of the dMCL and superficial anterior MCL and subsequent osteophyte removal, we achieved a satisfactory outcome. Posterior cruciate ligament (PCL) excision was routinely performed. We took care not to disturb the Pes Anserinus unless necessary. We did not find it necessary to release the posteromedial capsule, POL, or semimembranosus. In standard cases with moderate to mild varus deformities of 15° or less, which are the most common, satisfactory outcomes were achieved with the excision of the PCL along with the release of the dMCL and anterior sMCL. For such cases, multicenter studies could provide stepwise treatment algorithms that encompass fewer ligament releases and protect against excessive releases.

DECLARATIONS

Ethics Committee Approval: The study was approved by Metin Sabancı Baltalimanı Bone Diseases Training and Research Center Ethics Committee (No: 22/146, Date: 30/05/2024).

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: Not used.

Authorship Contributions: Concept – SG, HG; Design – HG; Supervision – HG; Fundings – SG, HB, NI, AK; Materials – SG, MBA, NI, AK; Data collection &/or processing – MBA; Analysis and/or interpretation – MBA; Literature search – SG, HB, AK, NI; Writing – SG; Critical review – HG.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Hutchison L, Grayson J, Hiller C, D'Souza N, Kobayashi S, Simic M. Relationship between knee biomechanics and pain in people with knee osteoarthritis: a systematic review and meta-analysis. Arthritis Care Res (Hoboken) 2023;75:1351–61.
2. Bercovy M, Kerboull L, Müller JH, Saffarini M, Sailhan F. Satisfactory mid- to long-term outcomes of TKA aligned using conventional instrumentation for flexion gap balancing with minimal soft tissue release. Knee Surg Sports Traumatol Arthrosc 2022;30:627–37.
3. Harris AB, Oni JK. Cementless, cruciate-retaining primary total knee arthroplasty using conventional instrumentation: technical pearls and intraoperative considerations. JBJS Essent Surg Tech 2024;14:e23.00036.
4. Sakai S, Nakamura S, Maeda T, Kuriyama S, Nishitani K, Morita Y, et al. Medial soft tissue release is also related to the anterior stability of cruciate-retaining total knee arthroplasty: a cadaveric study. Knee Surg Relat Res 2024;36:29.
5. Takashima Y, Takayama K, Ishida K, Muratsu H, Matsushita T, Niikura T, et al. Comparison of intraoperative soft tissue balance measurement between two tensor systems in total knee arthroplasty. Knee 2020;27:1071–7.
6. Nagai K, Muratsu H, Kanda Y, Tsubosaka M, Kamenaga T, Miya H, et al. Intraoperative soft tissue balance using novel medial preserving gap technique in posterior-stabilized total knee arthroplasty: comparison to measured resection technique. Knee Surg Sports Traumatol Arthrosc 2018;26:3474–81.
7. Kono K, Dorthe EW, Tomita T, Tanaka S, Angibaud L, D'Lima DD. Intraoperative knee kinematics measured by computer-assisted navigation and intraoperative ligament balance have the potential to predict postoperative knee kinematics. J Orthop Res 2022;40:1538–46.
8. Keggi JM, Wakelin EA, Koenig JA, Lawrence JM, Randall AL, Ponder CE, et al. Impact of intra-operative predictive ligament balance on post-operative balance and patient outcome in TKA: a prospective multicenter study. Arch Orthop Trauma Surg 2021;141:2165–74.
9. Batailler C, Hannouche D, Benazzo F, Parratte S. Concepts and techniques of a new robotically assisted technique for total knee arthroplasty: the ROSA knee system. Arch Orthop Trauma Surg 2021;141:2049–58.
10. Karasavvidis T, Pagan Moldenhauer CA, Haddad FS, Hirschmann MT, Pagnano MW, Vigdorchik JM. Current concepts in alignment in total knee arthroplasty. J Arthroplasty 2023;38(7 Suppl 2):S29–37.
11. Teeter MG, Perry KI, Yuan X, Howard JL, Lanting BA. Contact kinematic differences between gap balanced vs measured resection techniques for single radius posterior-stabilized total knee arthroplasty. J Arthroplasty 2017;32:1834–8.
12. Song SJ, Lee HW, Kim KI, Park CH. Load imbalances existed as determined by a sensor after conventional gap balancing with a tensiometer in total knee arthroplasty. Knee Surg Sports Traumatol Arthrosc 2020;28:2953–61
13. Song SJ, Kang SG, Lee YJ, Kim KI, Park CH. An intraoperative load sensor did not improve the early postoperative results of posterior-stabilized TKA for osteoarthritis with varus deformities. Knee Surg Sports Traumatol Arthrosc 2019;27:1671–9.
14. Geller JA, deMeireles AJ, Gazgalis A, Santos W, Neuwirth AL, Shah RP, et al. Collateral ligament tension and balance alone does not ensure improved outcome after total knee arthroplasty. J Arthroplasty 2023;38:S196–203.
15. Bardou-Jacquet J, Murgier J, Laudet F, Fabre T. Combining load sensor and robotic technologies for ligament balance in total knee arthroplasty. Orthop Traumatol Surg Res 2022;108:102889.
16. Neto JBA, Cavalcante MLC, Messias da Rocha PH, Helito CP, Lima LL, Ariel de Lima D. Study of the nerve endings and mechanoreceptors of the medial meniscotibial ligament of the knee: A structural and distribution analysis. Knee 2023;42:28–36.
17. Sakai S, Nakamura S, Maeda T, Kuriyama S, Nishitani K, Morita Y, et al. Medial soft tissue release is also related to the anterior stability of cruciate-retaining total knee arthroplasty: a cadaveric study. Knee Surg Relat Res 2024;36:29.
18. Bercovy M, Kerboull L, Müller JH, Saffarini M, Sailhan F. Satisfactory mid- to long-term outcomes of TKA aligned using conventional instrumentation for flexion gap balancing with minimal soft tissue release. Knee Surg Sports Traumatol Arthrosc 2022;30:627–37.
19. Javidan P, Ahmed M, Kaar SG. Arthroscopic release of the deep medial collateral ligament to assist in exposure of the medial tibiofemoral compartment. Arthrosc Tech 2014;3:e699–701.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2025.66934
Eur Arch Med Res 2026;42(2):118–123

A Comparative Analysis of Mini-Percutaneous Nephrolithotomy and Retrograde Intrarenal Surgery for the Management of 10–20 Mm Pediatric Kidney Stones

Recep Burak Degirmentepe,[1] Yasir Muhammed Akca,[1] Deniz Gul,[1] Haci Ibrahim Cimen,[1] Muammer Bozkurt,[2] Tuncay Toprak,[3] Fikret Halis[1]

[1]Department of Urology, Sakarya University, Faculty of Medicine, Sakarya, Turkiye
[2]Department of Urology, Health Sciences University, Prof. Dr. Cemil Tascioglu City Hospital, Istanbul, Turkiye
[3]Department of Urology, University of Health Sciences, Fatih Sultan Mehmet Training and Research Hospital, Istanbul, Turkiye

ABSTRACT

Objective: This study compares the clinical outcomes of mini-percutaneous nephrolithotomy (m-PNL) and retrograde intrarenal surgery (RIRS) in managing 10–20 mm renal stones in pediatric patients. The primary objective is to evaluate stone-free rates (SFRs), operative parameters, and complications to provide insights into optimal treatment strategies for pediatric nephrolithiasis.

Materials and Methods: A retrospective analysis of 39 pediatric patients under 18 years of age who underwent m-PNL (Group 1) or RIRS (Group 2) for renal stones between 10 and 20 mm from January 2016 to March 2024 was conducted. Patients were assessed for demographic data, operative time, hospitalization duration, fluoroscopy time, hemoglobin reduction, SFRs, and complications. Statistical analyses were performed to compare the two groups.

Results: The mean operative time, hospitalization duration, and fluoroscopy time were significantly shorter in the RIRS group ($p<0.001$), with lower post-operative hemoglobin reduction compared to the m-PNL group. At the 1-month follow-up, SFRs were comparable between groups (m-PNL: 95.2%, RIRS: 94.9%, $p=0.710$). Minor complications were observed in both groups without significant differences. RIRS demonstrated advantages in hospital stay and reduced radiation exposure, while m-PNL required fewer anesthesia sessions.

Conclusion: Both m-PNL and RIRS are effective and safe for treating pediatric kidney stones. While RIRS offers shorter operative and hospitalization times with reduced hemoglobin loss, m-PNL avoids the need for routine double-J stenting and additional anesthesia. Treatment choice should be individualized, considering clinical factors and family preferences, to optimize outcomes in pediatric nephrolithiasis management.

Keywords: Mini-percutaneous nephrolithotomy, Pediatric nephrolithiasis, Retrograde intrarenal surgery

Cite this article as: Degirmentepe RB, Akca YM, Gul D, Cimen HI, Bozkurt M, Toprak T, et al. A Comparative Analysis of Mini-Percutaneous Nephrolithotomy and Retrograde Intrarenal Surgery for the Management of 10–20 Mm Pediatric Kidney Stones. Eur Arch Med Res 2026;42(2):118–123.

Address for correspondence: Recep Burak Degirmentepe. Department of Urology, Sakarya University, Faculty of Medicine, Sakarya, Türkiye
E-mail: burakdegirmentepe@gmail.com **ORCID ID:** 0000-0002-2875-5750

Submitted: 10.07.2025 **Revised:** 02.08.2025 **Accepted:** 20.08.2025 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

INTRODUCTION

Pediatric nephrolithiasis is an infrequent condition; however, its incidence has increased significantly over the past few decades, with a dramatic annual rise of approximately 10%.[1] Managing renal stones in children presents greater challenges compared to adults, encompassing complexities in diagnosis, metabolic evaluations, surgical interventions, and prognosis.[2] To address this significant health issue, technological advancements in endourology have progressed rapidly over the years, aiming to achieve maximum success with minimal morbidity.[3]

Percutaneous nephrolithotomy (PNL) was first performed in the pediatric population in 1985 on a 14-year-old child.[3] PNL offers significantly high stone-free rates (SFRs) and requires fewer auxiliary procedures and sessions compared to other treatment methods. However, its application in children poses challenges due to the smaller size of their kidneys and collecting systems, the fragility of renal parenchyma, and the increased mobility of the pediatric kidney. Over time, the introduction of smaller instruments, as utilized in mini-PNL (m-PNL), has led to reduced parenchymal and vascular damage, making the procedure more suitable for this population.[4]

Currently, retrograde ıntrarenal surgery (RIRS) has emerged as a less invasive option to PNL for the treatment of renal calculi and has been performed with safety and efficacy. However, depending on the location and size of the stones, RIRS may necessitate multiple sessions.[3,5] With advancements in technology and increasing experience, both m-PNL and RIRS have demonstrated comparable success rates in achieving stone clearance.[3,6] Ideally, pediatric kidney stone treatments should aim for the highest SFRs with minimal complications, while also striving to minimize the number of surgical and anesthesia sessions required.[4,5]

We present a comparative clinical analysis of m-PNL and RIRS in the management of 10–20 mm renal stones in pediatric patients.

MATERIALS AND METHODS

A retrospective analysis was conducted on adolescents and children (<18 years) undergoing m-PNL or RIRS for kidney stones (10–20 mm) over an 8-year period (2016–2024). All patients under the age of 18 with 10–20 mm kidney stones who underwent m-PNL or RIRS during the specified period were selected for evaluation. Participants were stratified into two distinct cohorts based on the minimally invasive surgical technique employed: Group 1 (m-PNL) and Group 2 (RIRS). Demographic and clinical parameters, including gender, stone size, age, stone location, duration of hospitalization, operative time, fluoroscopy time, hemoglobin reduction, complications, and SFRs, were retrospectively reviewed, analyzed, and compared.

The study was conducted retrospectively, and the choice of surgical technique was not randomized. Operative management was standardized through single-surgeon performance for both techniques. The patients included in the study did not have any pathology that could alter the success of treatment, such as urinary obstruction, bleeding diathesis, ectopic kidney, malrotation, or ureteropelvic junction obstruction. All patients included in the study had unilateral renal stones. All patients' parents were given detailed information about kidney stone treatments before surgery. Ethical clearance was granted by Sakarya University's Institutional Review Board (No: 123, Date: 15/04/2024). The study was conducted in accordance with the principles of the Declaration of Helsinki.

Preoperative evaluations included routine biochemical and microbiological testing. All procedures were conducted only after confirmation of negative urine cultures. Patients with positive urine cultures were treated with targeted antibiotics. Radiological assessments, as necessary, were performed using renal ultrasonography (US), intravenous urography, or computed tomography (CT). Stone size was defined as the longest axis of each stone measured using preoperative US, kidney-ureter-bladder (KUB) radiography, or CT. Stone size was defined as the longest axis of each stone based on preoperative imaging. Due to the retrospective design and variation in imaging modalities over the years, consistent data on the second axis were not available. Therefore, volume or surface area calculations could not be applied uniformly across all cases, which is a noted limitation of the study. In cases of multiple stones, the total stone burden was calculated by summing the longest axes of each individual stone.

Intravenous cephalosporin was administered to all patients as prophylaxis before the endoscopic procedure, with the dosage adjusted based on age and weight. For m-PNL, patients were initially positioned in the dorsal lithotomy position, where a ureteral catheter was inserted over a guidewire to the targeted kidney. The external portion of the catheter was secured to the inserted urethral catheter using a silk suture. Patients were then repositioned to the prone position, and the kidney was visualized using a C-arm fluoroscope by administering contrast material through the pre-inserted ureteral catheter. Access to the renal pelvicalyceal system was achieved through an 18-gauge Chiba needle inserted into the appropriate calyx, followed by guidewire advancement into the renal pelvis and calyces. Tract dilation was performed over the guidewire, and an access sheath was positioned. A 12Fr nephroscope (Richard Wolf, Knittlingen, Germany) was introduced through a 15/18 Fr sheath to access the pyelocalyceal system. Calculi were disintegrated utilizing a 550-micron holmium: YAG laser fiber and pneumatic lithotripters, and the resulting fragments were retrieved with forceps. Upon achieving stone clearance, a nephrostomy catheter was placed in all m-PNL patients, clamped on post-operative day 1, and removed on day 2. The total operation time for m-PNL

encompassed patient positioning, ureteral catheter placement through cystoscopy, percutaneous access, tract dilation, stone fragmentation, and nephrostomy placement.

The RIRS procedure involved general anesthesia, lithotomy positioning, and fluoroscopically-guided insertion of a 9.5 Fr ureteral access sheath (28/35 cm, Cook Medical) through which a 7.5 Fr flexible ureterorenoscope (Flex X2, Karl Storz) was advanced. After the stone was once identified, the stone was fragmented into smaller pieces to facilitate spontaneous passage, utilizing a 270 µm fiber holmium:YAG laser set at 0.6 J and 6–8 Hz. Following stone fragmentation, a guidewire was advanced into the renal collecting system through the UAS. A double-J (DJ) stent was routinely placed under cystoscopic guidance over the guidewire in all patients. This stent was scheduled for removal under general anesthesia 4-week post-procedure. Preoperative stenting was not routinely performed before RIRS procedures. The total operation time for RIRS was defined as the duration encompassing the induction of general anesthesia, patient positioning, cystoscopic placement of a guidewire into the ureter, UAS placement, stone fragmentation, and DJ stent placement at the end of the procedure.

Post-operative follow-up was conducted using KUB radiography and US. The patients were routinely checked for residual stones with KUB radiography on the 1st day post-operatively and KUB radiography and US on the 1st month post-operatively. Treatment success was objectively defined as complete stone-free status, confirmed by imaging modalities at the 1-month post-operative follow-up. Post-operative complications were evaluated for each patient using the Clavien–Dindo classification system.[7] Treatment efficacy was principally assessed through SFR, where success mandated the absence of any radiologically detectable stone fragments larger than 4 mm. Secondary analyses focused on procedure-related adverse events, which were prospectively documented and severity-stratified according to Clavien–Dindo classifications (Grades I-V).

Statistical Analysis

All data processing utilized the Statistical Package for the Social Sciences 16.0J software (IBM). Nonparametric approaches (Mann–Whitney U for continuous data; χ^2 tests for proportions) were selected due to small cohort size and distributional concerns. The alpha threshold for significance was established at 0.05.

RESULTS

The mean age of the patients was 7.6±2.8 years in the m-PNL group and 8.2±2.7 years in the RIRS group (p=0.600). Group 1 included 12 males (57%) and 9 females (43%), while Group 2 comprised 10 males (55%) and 8 females (45%). The mean stone size was 13.9±0.9 mm in Group 1 and 13.6±1.2 mm in Group 2 (p=0.349), respectively. The demographic data of the patients are summarized in Table 1.

Table 1. Demographic values of the patients

	Mini-PNL	RIRS	p
Number	21	18	
Male/female	12/9	10/8	
Age (y)*	7.6±2.8	8.2±2.7	0.600
Stone size (mm)*	13.9±0.9	13.6±1.2	0.349
Stone locations**			
Lower	11	3	
Middle	2	0	
Upper	1	7	
Pelvic	5	7	
Multicalyceal	2	1	

PNL: Percutaneous nephrolithotomy; RIRS: Retrograde intrarenal surgery; *Mean±standard deviation; **Distribution of stones in the kidney.

The mean operation time was 66.8±10.6 min for Group 1 and 49.2±7 min for Group 2, showing a statistically significant difference (p<0.001). The mean hospitalization time was shorter in the RIRS group (26.3±4.6 h) compared to the m-PNL group (77.7±18.4 h), as expected with a less invasive procedure (p<0.001). Fluoroscopy exposure during surgery averaged 138.6±19.5 s for Group 1 and 30.4±28.1 s for Group 2, with the difference again being statistically significant (p<0.001).

The post-operative hemoglobin decrease was 1.65±0.31 g/dL in the m-PNL group and 0.5±0.29 g/dL in the RIRS group, with a statistically significant greater decrease observed in the m-PNL group. Stone location distribution differed significantly between groups: Lower pole stones predominated in the m-PNL group (52.3% vs. 16.6% in RIRS), while upper pole stones were more frequent in the RIRS group (38.8% vs. 4.7% in m-PNL). The SFRs on the 1st post-operative day were 95.2% for the m-PNL group and 88.9% for the RIRS group, with no statistically significant difference (p=0.630). In the m-PNL group, two patients had multicalyceal stones. One patient was managed with a single lower pole access, while the second required two access tracts due to anatomical constraints. Both cases achieved stone-free status postoperatively. One RIRS case (5.6%) required staged treatment due to failed initial UAS placement. After deferred DJ stenting and a 3-week interval, UAS placement and stone fragmentation were successfully completed. At the 1-month follow-up, the SFRs were 95.2% for the m-PNL group and 94.9% for the RIRS group, with no significant difference between the groups (p=0.710).

No major complications were observed in the post-operative period. In the m-PNL group, a DJ stent was placed in one patient due to symptomatic ureteral stones detected during

Table 2. Comparison of post-operative outcomes

	Mini-PNL	RIRS	p
Operation time* (min)	66.8±10.6	49.2±7	<0.001
Hospitalization* (h)	76.7±18.4	26.3±4.6	<0.001
Fluoroscopy time* (s)	138.6±19.5	30.4±28.1	<0.001
Hemoglobin drop* (g/dL)	1.65±0.31	0.5±0.29	<0.001
Stone-free rate,** n (%)	19/21 (90.5)	16/18 (88.9)	0.630
Success rate,*** n (%)	20/21 (95.2)	17/18 (94.9)	0.710
Post-op complications			0.551
Clavien grade I	5	2	
Clavien grade II	2	2	
Clavien grade III	1	0	
Clavien grade IV-V	0	0	

PNL: Percutaneous nephrolithotomy; RIRS: Retrograde intrarenal surgery; *Mean±standard deviation; **Outcomes on post-operative day 1; ***Stone-free rate at 1st month.

follow-up. No cases of bleeding requiring blood transfusion occurred. Minor post-operative complications, such as hematuria, urinary infection, ecchymosis, and paralytic ileus, were observed in some patients in either group and were managed appropriately. The comparison of complications between the groups, according to the Clavien–Dindo classification, is summarized in Table 2.

DISCUSSION

Technological advancements in urology have revolutionized therapeutic approaches for nephrolithiasis, replacing open surgeries with minimally invasive techniques such as Extracorporeal Shock Wave Lithotripsy (ESWL), PNL, and RIRS.[2] ESWL is particularly effective for 10–20 mm kidney stones in selected cases; however, its success rate declines as stone size increases. Achieving higher SFRs often requires multiple sessions, leading to additional anesthesia and radiation exposure.[8] The miniaturization of instruments and the development of flexible devices have facilitated easier access to renal stones, thereby expanding the applicability of m-PNL and RIRS procedures in the pediatric population.[2]

Compared to adults, pediatric kidneys are more mobile, have a narrower pelvicalyceal system, and possess more fragile renal parenchyma and mucosa, making it challenging to balance complete stone clearance with the preservation of renal parenchyma.[9] The primary goal in pediatric urolithiasis treatment is to employ a minimally invasive surgical technique that minimizes impact on kidney function while achieving a high stone clearance rate on the first attempt, without the need for re-treatment or auxiliary procedures.[10]

The rising prevalence of pediatric nephrolithiasis in recent years highlights the need for a thorough assessment of surgical techniques employed in its treatment.[11] Accordingly, we aimed to contribute to this field by presenting our clinical study, which compares m-PNL and RIRS in the pediatric patients.

There is comparatively limited research comparing m-PNL and RIRS in children. Resorlu et al.[6] conducted a multi-institutional analysis focused on kidney stones between 10 and 30 mm in size, reporting an overall SFR of 85.8% in the m-PNL group and 84.2% in the RIRS group after one session. Their analysis demonstrated significantly different success rates between modalities: RIRS achieved 87.3% SFR for stones <20 mm versus 50% for >20 mm calculi, whereas m-PNL showed superior efficacy (100% and 83.9% SFR for respective size categories). The final SFR was improved to 94.3% in the m-PNL group and 92.6% in the RIRS group. Although they found longer mean fluoroscopy time, operation time, and hospitalization duration in the m-PNL group, their findings indicated that the morbidities associated with PNL could be significantly reduced using the RIRS technique. In a similar study, Saad et al.[12] conducted a randomized controlled study comparing the outcomes of PNL and RIRS in 38 pediatric patients with kidney stones larger than 2 cm. They found 71% and 95.5% SFRs for RIRS and PNL, respectively. According to the study results, there were higher SFRs for PNL, less radiation exposure, shorter hospital stay, and fewer complications for RIRS. They suggested that RIRS was a preferred surgical option but not an innocent procedure. These findings, however, were applicable to larger stone sizes, while our study focused on stones ranging from 10–20 mm. In addition, while we used a 12Fr nephroscope in our study, Saad et al.[12] employed a 22Fr nephroscope, which further differentiates the two studies.

Recently, several researchers have compared m-PNL and RIRS as surgical techniques for treating kidney stones ranging from 10 to 20 mm in children. Wang et al.[13] investigated children under 3 years with solitary 10–20 mm stones, demonstrating both techniques as clinically effective, though m-PNL proved superior by requiring fewer anesthesia sessions. Corroborating these findings, Özçift et al.[4] reported comparable outcomes in their pediatric cohort (stone size 10–20 mm), where RIRS offered advantages in reduced hospitalization duration and shorter fluoroscopy exposure - aligning with our institutional experience. However, their cost-benefit analysis favored m-PNL for its economic efficiency and reduced anesthetic burden. In our study, which focused on pediatric kidney stones ranging from 10 to 20 mm, we found a statistically significant difference between the m-PNL and RIRS groups regarding operation time, hospitalization duration, fluoroscopy time, and hemoglobin decrease, all favoring RIRS. However, no significant differences were observed in terms of SFR, overall suc-

cess rate, or post-operative complications. Besides, our study revealed a disparity in stone location between groups, with lower pole stones more common in the m-PNL cohort and upper pole stones in the RIRS cohort. This imbalance may have influenced outcomes, as lower pole stones are technically more challenging to treat with RIRS due to difficulties in ureteroscope deflection and fragment clearance, while upper pole stones pose access challenges for percutaneous approaches. Despite this, both techniques achieved comparable SFRs, suggesting that surgeon expertise and instrument selection (e.g., flexible nephroscopes in m-PNL) may mitigate anatomical constraints. Future studies should stratify outcomes by stone location to control for this bias.

The ongoing Pediatric Kidney Stone Care Improvement in Children and Teens (PKIDS) study, a large comparative effectiveness trial in the United States, is further investigating the outcomes of PNL, ureteroscopy, and shock wave lithotripsy in over 1000 pediatric patients. This study aims to provide robust data on the optimal management of pediatric nephrolithiasis and may offer additional insights into the comparative efficacy of these techniques.[14] Our findings align with the preliminary trends observed in PKIDS, particularly in terms of the safety and efficacy of minimally invasive techniques like RIRS and m-PNL in children.

The neurodevelopmental consequences of anesthesia exposure in infants and young children remain incompletely understood, despite millions of pediatric procedures performed annually under general anesthesia. Preclinical studies demonstrate that common anesthetic agents induce dose-dependent neurotoxicity through mechanisms including apoptotic neurodegeneration, synaptic dysfunction, and impaired neurogenesis – particularly during peak synaptogenesis (6–24 months postnatally). While human cohort studies report conflicting associations with later cognitive deficits (e.g., learning disabilities), current evidence suggests that repeated or prolonged exposure may elevate neurodevelopmental risks, warranting judicious use of anesthesia in early childhood. This underscores the importance of technique selection (e.g., regional adjuncts), duration minimization, and adherence to evolving consensus guidelines for elective procedures in high-risk developmental windows.[15,16] The study by Ebert et al.[17] highlights the preference for spinal anesthesia as a strategy to mitigate the potential neurocognitive risks associated with general anesthesia in children, aligning with these findings. In addition, the study conducted by Jia et al.[15] further indicates that factors such as more frequent exposure to surgery and anesthesia, younger age at the time of exposure, and lower body weight may have detrimental effects on cognitive and behavioral development. It has been reported that repeated exposure to anesthesia may impact brain development in children, potentially affecting their intellectual growth and cognitive abilities.[13] While some studies suggest potential risks associated with repeated or prolonged exposure to general anesthesia, large population-based studies indicate that a single brief exposure is not linked to long-term neurodevelopmental deficits.[18] In our study, the comparable SFRs achieved by the RIRS group with a shorter operation time suggest that RIRS could be a valuable option for pediatric patients with 10–20 mm kidney stones.

Our study has several limitations. Its retrospective design introduces potential bias. The absence of randomization, the relatively small sample size in each group, and the lack of comprehensive stone analysis for all patients represent significant constraints. The heterogeneity in imaging modalities used for stone size measurement is another limitation of our study. While we prioritized CT measurements for accuracy, variations in imaging techniques may have introduced some bias. Future studies should standardize imaging protocols to ensure more consistent and reliable comparisons. Another important limitation of our study is the measurement of stone size based solely on the longest axis. This approach may not accurately reflect the actual stone burden, particularly in cases of irregularly shaped or elongated stones. Due to the retrospective design and variability in imaging methods, consistent measurements of the second axis or volumetric assessments could not be performed. Furthermore, data regarding stone hardness, including Hounsfield Unit values or compositional analysis, were not consistently available and could not be incorporated into our analysis. Given the potential influence of stone hardness on fragmentation and procedural efficiency (especially in RIRS), this represents an additional limitation. An additional limitation is that the non-randomized design led to uneven distribution of stone locations between groups, potentially favoring m-PNL for lower pole stones and RIRS for upper pole stones. While this reflects real-world clinical decision-making, it may have confounded operative outcomes. This imbalance may have introduced a selection bias, as anatomical location can influence both access and fragmentation efficiency depending on the surgical approach. Future prospective, randomized studies with larger patient cohorts are necessary to obtain more robust and meaningful findings. On the other hand, the fact that all surgical procedures were performed by a single surgeon is a strength of our study, as it ensures consistency in surgical technique and minimizes variability in outcomes.

CONCLUSION

M-PNL and RIRS are endoscopic surgical techniques proven to be effective and safe for pediatric kidney stone management. RIRS has demonstrated advantages such as shorter hospital stays and operation times, reduced post-operative hemoglobin loss compared to m-PNL, and lower fluoroscopy expo-

sure during surgery. However, despite achieving comparable stone-free and success rates with m-PNL, the routine use of DJ stents after RIRS, the need for additional anesthesia during stent removal, and the potential requirement for further minimally invasive procedures should be carefully considered. These factors underscore the importance of individualized treatment decisions, particularly in pediatric patients, with the preferences of the child's family taken into account.

DECLARATIONS

Ethics Committee Approval: The study was approved by Sakarya University Ethics Committee (No: 123, Date: 15/04/2024).

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: Not declared.

Authorship Contributions: Concept – RBD, YMA, DG, HIC, MB, TT, FH; Design – RBD, YMA, DG, HIC, MB, TT, FH; Supervision – RBD, YMA, DG, HIC, MB, TT, FH; Fundings – FH; Materials – MB; Data collection &/or processing – DG, TT; Analysis and/or interpretation – RBD; Literature search – YMA; Writing – RBD; Critical review – HIC.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Baum M. Editorial: Pediatric nephrolithiasis. Curr Opin Pediatr 2020;32:261–4.
2. Zhao FZ, Li J, Tang L, Li CM, Zhang Y, Wang WY, et al. Comparison of efficacy and safety of minimally invasive procedures for 10-20 mm pediatric renal Stones-A bayesian network meta-analysis. J Pediatr Urol 2020;16:771–81.
3. Karatag T, Tepeler A, Silay MS, Bodakci MN, Buldu I, Daggulli M, et al. A Comparison of 2 percutaneous nephrolithotomy techniques for the treatment of pediatric kidney stones of sizes 10-20 mm: microperc vs miniperc. Urology 2015;85:1015–8.
4. Özçift B, Tiryaki HT. Comparison of miniaturized percutaneous nephrolithotomy and retrograde intrarenal surgery: Which is more effective for 10-20 mm renal stones in children? Arch Esp Urol 2020;73:643–50. [Article in English, Spanish]
5. Türk C, Petřík A, Sarica K, Seitz C, Skolarikos A, Straub M, et al. EAU guidelines on interventional treatment for urolithiasis. Eur Urol 2016;69:475–82.
6. Resorlu B, Unsal A, Tepeler A, Atis G, Tokatli Z, Oztuna D, et al. Comparison of retrograde intrarenal surgery and mini-percutaneous nephrolithotomy in children with moderate-size kidney stones: results of multi-institutional analysis. Urology 2012;80:519–23.
7. Bolliger M, Kroehnert JA, Molineus F, Kandioler D, Schindl M, Riss P. Experiences with the standardized classification of surgical complications (Clavien-Dindo) in general surgery patients. Eur Surg 2018;50:256–61.
8. Smaldone MC, Gayed BA, Ost MC. The evolution of the endourologic management of pediatric stone disease. Indian J Urol 2009;25:302–11.
9. Chen Y, Deng T, Duan X, Zhu W, Zeng G. Percutaneous nephrolithotomy versus retrograde intrarenal surgery for pediatric patients with upper urinary stones: a systematic review and meta-analysis. Urolithiasis 2019;47:189–99.
10. Samotyjek J, Jurkiewicz B, Krupa A. Surgical treatment methods of urolithiasis in the pediatric population. Dev Period Med 2018;22:88–93.
11. Miah T, Kamat D. Pediatric nephrolithiasis: a review. Pediatr Ann 2017;46:e242–4.
12. Saad KS, Youssif ME, Al Islam Nafis Hamdy S, Fahmy A, El Din Hanno AG, El-Nahas AR. Percutaneous nephrolithotomy vs retrograde intrarenal surgery for large renal stones in pediatric patients: a randomized controlled trial. J Urol 2015;194:1716–20.
13. Wang W, Ge Y, Wang Z, Wang L, Li J, Tian Y. Comparing micropercutaneous nephrolithotomy and retrograde intrarenal surgery in treating 1-2 cm solitary renal stones in pediatric patients younger than 3 years. J Pediatr Urol 2019;15:517.e1–e6.
14. Ellison JS, Lorenzo M, Beck H, Beck R, Chu DI, Forrest C, et al. Comparative effectiveness of paediatric kidney stone surgery (the PKIDS trial): study protocol for a patient-centred pragmatic clinical trial. BMJ Open 2022;12:e056789.
15. Jia X, Tan S, Qin Y, Wei Y, Jiang Y, Pan S, et al. Experiencing anesthesia and surgery early in life impairs cognitive and behavioral development. Front Neurosci 2024;18:1406172.
16. Xu J, Mathena RP, Singh S, Kim J, Long JJ, Li Q, et al. Early developmental exposure to repetitive long duration of midazolam sedation causes behavioral and synaptic alterations in a rodent model of neurodevelopment. J Neurosurg Anesthesiol 2019;31:151–62.
17. Ebert KM, Jayanthi VR, Alpert SA, Ching CB, DaJusta DG, Fuchs ME, et al. Benefits of spinal anesthesia for urologic surgery in the youngest of patients. J Pediatr Urol 2019;15:49.e1–e5.
18. Liu X, Ji J, Zhao GQ. General anesthesia affecting on developing brain: evidence from animal to clinical research. J Anesth 2020;34:765–72.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2025.26023
Eur Arch Med Res 2026;42(2):124–132

Evaluation of Attitudes, Knowledge Levels, and Perceived Quality of Nutritional Care Among Nurses Working in Internal Medicine and Intensive Care Clinics: A Descriptive Cross-Sectional Study

Nese Iscan Ayyildiz,[1] Sevgi Demir Cam[2,*]

[1]Department of Health Care Services, Trabzon University, Tonya Vocational School, Trabzon, Turkiye
[2]Department of Nursing, Artvin Coruh University, Faculty of Health Sciences, Artvin, Turkiye
The current affiliation of the author: Department of Nursing, Hitit University, Faculty of Health Sciences, Çorum, Türkiye

ABSTRACT

Objective: Malnutrition is an important complication that is common in critically ill patients and affects the efficiency of clinical treatment. The level of knowledge and attitudes of nurses, who have an important role in clinical treatment, about nutritional care, has an important place in patient care. Various factors affect the knowledge of and attitudes to nutritional care. The aim of this study was to evaluate the nutritional attitudes, levels of knowledge, and perceptions toward care in nurses working in internal medicine and intensive care clinics.

Materials and Methods: The study was completed with 265 internal medicine and intensive care clinic nurses who volunteered to participate in the study and were not on leave during the data collection process without sample selection.

Results: It was stated by 53.6% of the nurses that they had not received nutritional education, 59.62% reported a moderate level of nutritional care knowledge, and 56.6% were willing to receive nutritional education. The mean scores of the nurses were found to be 22.95±3.1 for attitudes toward the importance of nutritional assessment, 26.77±3.53 for level of knowledge of nutritional care, and 33.63±5.12 for perceived quality of nutritional care.

Conclusion: The nurses in this study had positive attitudes toward the importance of nutritional assessment and a moderate level of knowledge of nutritional care and perceived quality of nutritional care.

Keywords: Attitude, Nursing, Nutritional assessment, Nutritional care, Perception

Cite this article as: Iscan Ayyildiz N, Demir Cam S. Evaluation of Attitudes, Knowledge Levels, and Perceived Quality of Nutritional Care Among Nurses Working in Internal Medicine and Intensive Care Clinics: A Descriptive Cross-Sectional Study. Eur Arch Med Res 2026;42(2):124–132.

INTRODUCTION

Nutrition is a fundamental requirement of human life and represents a modifiable risk factor both for maintaining health and for preventing disease.[1] Patients may be at risk of poor nutrition before or after hospitalization, which can lead to malnutrition.[2,3] The incidence of malnutrition is reported as 22% among hospitalized patients and at least 50% in intensive care patients.[2] Despite its critical role in disease management and recovery, nutrition is sometimes neglected in clinical practice.[4-6]

Address for correspondence: Nese Iscan Ayyildiz. Department of Health Care Services, Trabzon University, Tonya Vocational School, Trabzon, Türkiye

E-mail: ayyildiz2861@gmail.com **ORCID ID:** 0000-0003-0976-8000

Submitted: 01.07.2025 **Revised:** 14.08.2025 **Accepted:** 15.08.2025 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Preventing malnutrition-related complications requires early identification of at-risk patients, continuous monitoring, and timely nutritional support.[5,7] This is particularly important in intensive care settings, where malnutrition negatively affects patient safety and quality of care. Current guidelines recommend that patients admitted to intensive care and clinical wards should be screened for malnutrition risk and provided with appropriate nutritional interventions within 24 h of admission.[8,9]

Nutritional care is delivered by an interdisciplinary team involving physicians, nurses, pharmacists, and dietitians. Among these, nurses play a pivotal role: They are responsible for screening nutritional status, initiating and maintaining nutritional support, recognizing complications, and delivering evidence-based nursing interventions.[7,10,11] Clinical nurses are expected to identify high-risk groups, monitor patient changes during nutritional therapy, and take prompt action when complications occur.[12] This requires sufficient awareness, knowledge, and skills regarding nutritional care, particularly among those working in intensive care and internal medicine clinics.[13]

Previous studies have evaluated nurses' knowledge, attitudes, and perceived quality of nutritional care in various settings,[1,14] yet there is limited evidence focusing specifically on internal medicine and intensive care nurses in Turkey or globally. Understanding nurses' perceptions and practices in these critical areas is essential for designing effective training programs and improving patient outcomes. This study aims to address this gap by assessing the attitudes, knowledge levels, and perceived quality of nutritional care among nurses working in internal and intensive care clinics. We hypothesize that nurses working in internal medicine and intensive care units demonstrate a generally positive attitude toward nutritional assessment, possess moderate to high levels of knowledge, and perceive the quality of nutritional care in their units as adequate, although certain areas may require improvement through continuous training and institutional support.

MATERIALS AND METHODS

Study Design

This descriptive and cross-sectional study was conducted with nurses working in the internal medicine and intensive care clinics of two district hospitals in northern Turkey between September 05, 2024, and January 05, 2025.

Setting and Sample

The study universe consisted of a total of 271 nurses working in the internal medicine and intensive care clinics of two district hospitals. No selection method was applied and 268 internal and intensive care clinic nurses who volunteered to participate and were not on leave/sick leave during the data collection process were reached; 3 nurses were not included in the study due to incomplete information in the data collection forms and the study was completed with a total sample of 265 nurses.

Inclusion Criteria

1. Age >18 years
2. Ability to communicate verbally
3. Actively working in internal medicine clinics and intensive care units for at least 6 months
4. No physical or mental health problems that would prevent comprehension and the ability to respond to the questions
5. Voluntary participation in the study.

After the purpose of the study was explained, written and verbal consent was obtained from all individuals participating in the study. The research data were collected using the survey method and face-to-face interview technique.

Data Collection Instruments

Data were collected using a Nurse Personal Information Form and an Assessment Questionnaire evaluating the importance of nutritional evaluation, nurses' knowledge level, and perceived quality of nutritional care.

Nurse Information Form

This form, which was prepared in line with the literature,[5] includes 10 questions about age, gender, educational status, marital status, years of service in the profession, the clinic where the nurses work, years of service in the clinic where they work, the status of receiving education on nutritional support, the level of knowledge they perceive about nutritional care, and their willingness to receive education on nutritional care.

Evaluation Questionnaire of The Significance of Nutritional Evaluation, Nurses' Knowledge, and Their Perceived Quality of Nutritional Care

This questionnaire was developed by Miriam Theilla et al.[13] and validity and reliability studies of the Turkish version were conducted by Gürlek Kısacık et al.[5] It consists of three sections; the first section has seven items to assess the attitudes of nurses toward the clinical importance of nutritional evaluation and their perceptions of nutritional assessment as a basic component of nursing care. The total score of this section ranges from 7 to 28 points, with a higher score indicating a more positive outlook toward the importance of assessing the nutritional status of patients and acceptance of nutritional assessment among nursing functions.

The second section of the questionnaire consists of 10 items evaluating the knowledge level of nurses regarding nutritional care. These items are reverse scored to give a total in the range of 10–40 points, with a higher score demonstrating a higher level of knowledge of nutritional care. The third section consists of nine items evaluating how nurses perceive the quality of nutritional care provided to patients in the clinics where they work. The total score of this section ranges from 9 to 45 points with a higher score indicating that nurses evaluate the quality of nutritional care provided to patients in the units where they work more positively. The items in the questionnaire are scored on a 4-point Likert scale as "strongly disagree" (1), "disagree" (2), "agree" (3), "strongly agree" (4). In the Turkish validity and reliability study, Gürlek Kısacık et al.[5] found Cronbach's alpha coefficient to be α=0.82 for the attitude factor, α=0.63 for the level of knowledge factor, and α=0.85 for the perceived quality of care factor. In the present study, the Cronbach's alpha coefficient was determined to be 0.83 for attitude regarding the importance of nutritional assessment, 0.71 for level of knowledge regarding nutritional care, and 0.87 for perceived quality of nutritional care.

Statistical Analysis

Descriptive statistics were stated as mean±standard deviation values for continuous variables and as number (n) and percentage (%) for categorical variables. The conformity of numerical variables to the normal distribution was evaluated with Skewness and Kurtosis. Data that showed normal distribution were examined with the Independent t-test, one-way ANOVA, and correlation analysis. As the distribution of variances was homogeneous, "Least Significant Difference" was used as further analysis. A value of $p<0.05$ was considered statistically significant.

Ethical and Institutional Approval

All participants were fully informed about the study and all provided informed consent for participation. The research received ethical approval from the Trabzon University Social and Human Sciences Scientific Research and Publication Ethics Committee (E-81614018-050.04-2400014629/18.03.2024), and institutional permission was granted by the two district hospitals (E-55568733-604.01-245799120/June 07, 2024). Every phase of the study was conducted in compliance with the Helsinki Declaration.

RESULTS

Participant Characteristics

Evaluation was made of a total of 265 nurses comprising 72.5% females and 27.5% males with a mean age of 40.02±7.97 years, 85.7% were married, and 71.7% were undergraduates. The duration of working as a nurse was ≥20 years in 40.8% of the participants, 33.6% were working in the internal medicine clinic, and 47.9% had been working in the clinic for 5–10 years. It was stated by 53.6% of the participants that they had not received any nutritional education, 59.62% stated that their level of nutritional care knowledge was at an intermediate level, and 56.6% stated that they wanted to receive nutritional education (Table 1).

The Attitude of Nurses Regarding the Importance of Nutritional Evaluation, the Level of Knowledge about Nutritional Care, and the Factors Affecting the Perception of Nutritional Care Quality

No significant difference was found between the gender and marital status of the participants and the mean scores of attitudes regarding the importance attributed to nutritional evaluation, their level of knowledge of nutritional care, and their perception of care quality ($p>0.05$) (Table 1).

A statistically significant difference was determined in the mean scores of attitudes with respect to the perceived importance of nutritional evaluation and level of knowledge of nutritional care according to the educational status of the participants and it was found that participants with an undergraduate degree had higher mean scores ($p<0.05$). No notable variation was observed in perceived quality of nutritional care across different educational statuses ($p>0.05$) (Table 1).

A statistically significant difference was found between the years of service and the mean scores of attitudes regarding the significance of nutritional evaluation, the extent of knowledge of nutritional care, and the perceived standard of nutritional care. Nurses who had worked for 20 years and more were found to have higher scores ($p<0.05$) (Table 1).

A statistically significant difference was found between the mean scores of perceived quality of nutritional care according to the clinic in which the nurses were working. Nurses who worked in the palliative care unit, intensive care unit, and chest diseases clinic were seen to have higher scores than the nurses who worked in cardiology and neurology clinics ($p<0.05$). No significant difference was observed between the mean scores of attitudes regarding the significance of nutritional evaluation and the extent of knowledge about nutritional care according to the clinics where they worked ($p>0.05$) (Table 1).

A significant difference was observed in the mean scores of attitudes regarding the perceived importance of nutritional evaluation, knowledge level in nutritional care, and perceived quality of that care according to the duration of working in the clinic. Nurses who had worked for 20 years or more were found to have higher average scores related to the perceived importance of nutritional assessment. It was also seen that those who had worked for 1–5 years and 5–10 years had higher level of knowledge of nutritional care mean scores than the nurses who had worked for 20 years and more and that the nurses

Table 1. Distribution of the mean scores of attitudes toward nutritional evaluation, un-derstanding of nutritional care, and perceived care quality according to the sociodemo-graphic characteristics of the participants

	n	Percentage	Attitudes toward the Importance of Nutritional Evaluation X±SD	Level of Knowledge of Nutritional Care X±SD	Perceived Quality of Nutritional Care X±SD
Gender					
Female	192	72.5	22.84±3.22	26.55±3.53	33.98±5.25
Male	73	27.5	23.24±2.81	27.34±3.47	32.69±4.66
Test and Significance			t: –0.940	t:–1.621	t:1.841
			p:0.348	p:0.106	p:0.067
Marital status					
Married	227	85.7	22.91±3.17	26.79±3.35	33.62±4.74
Single	38	14.3	23.18±2.76	26.65±4.49	33.71±7.05
Test and significance			t:–0.490	t:0.218	t: –0.075
			p:0.625	p:0.828	p:0.940
Educational status					
High school	7	2.6	20.57±2.43a	26.14±2.67a	32.28±2.98
Associate degree	68	25.7	21.92±2.97b	25.88±2.35b	33.89±4.77
Undergraduate degree	190	71.7	23.41±3.07c	27.11±3.85c	33.58±5.31
Test and Significance			F:8.214	F:3.218	F:0.337
			p:0.000	p:0.042	p:0.714
			c>a=b	c>b	
Years of service					
0–1 Years	2	0.8	20.50±0.70a	28.50±2.12a	29.00±5.65a
1–5 Years	31	11.7	23.58±2.83b	26.54±4.21b	33.64±6.92b
5–10 Years	33	12.5	24.21±2.91c	28.24±3.83c	33.87±5.02c
10–15 Years	50	18.7	23.78±2.91d	28.12±3.48d	31.66±4.48d
15–20 Years	41	15.5	22.63±3.14e	26.78±2.97e	33.12±4.84e
20 years and more	108	40.8	22.17±3.14f	25.73±3.13f	33.12±4.84f
Test and significance			F:3.919	F:4.934	F:3.047
			p:0.002	p:0.000	p:0.011
			c>e>f	c=d>b c=d>f	f>d
Clinics where nurses worked					
Internal medi-cine	89	33.6	22.95±3.42	26.35±3.85	33.35±5.09a
Palliative care	45	17.0	23.66±3.12	27.64±2.29	34.53±6.46b
Intensive care	60	22.6	23.30±2.81	27.23±3.84	34.70±4.57c
Cardiology	43	16.2	22.18±3.08	26.51±3.78	31.90±4.31d
Neurology	21	7.9	21.90±2.68	26.76±2.73	31.76±3.76e
Chest diseases	7	2.7	23.28±1.38	24.14±1.86	38.42±2.14f
Test and significance			F:1.655	F:1.847	F:3.803
			p:0.146	p:0.104	p:0.002
					b=c=f>d=e

Table 1. Distribution of the mean scores of attitudes toward nutritional evaluation, un-derstanding of nutritional care, and perceived care quality according to the sociodemo-graphic characteristics of the participants (Cont.)

	n	Percentage	Attitudes toward the Importance of Nutritional Evaluation X±SD	Level of Knowledge of Nutritional Care X±SD	Perceived Quality of Nutritional Care X±SD
Years of service in the clinic					
0–1 Years	3	1.1	20.66±0.57a	27.66±2.08a	30.00±4.35a
1–5 Years	58	21.9	23.86±2.83b	26.93±3.87b	34.06±6.70b
5–10 Years	127	47.9	23.10±3.09c	27.51±3.44c	32.77±4.78c
10–15 Years	66	23.9	21.65±3.07d	25.54±2.86d	32.60±1.14d
15–20 Years	5	1.9	23.30±1.78e	28.20±3.56e	36.50±2.81e
20 years and more	6	2.3	26.33±1.96f	21.50±0.83f	36.50±2.81f
Test and significance			F:5.331	F:6.146	F:2.345
			p:0.000	p:0.000	p:0.042
			f>c>d	b>d>e, c>d>f	d>c
The status of receiving nutritional education					
Yes	142	53.6	24.35±2.65	27.66±3.96	34.35±5.22
No	123	46.4	21.33±2.81	25.73±2.60	32.79±4.89
Test and significance			t:9.003	t:4.732	t:2.500
			p:0.000	p:0.000	p:0.000
Level of knowledge of nutritional care					
Insufficient	10	3.8	21.10±2.60a	26.50±2.87	30.10±5.06a
Moderate	158	59.6	22.48±3.11b	26.47±3.57	33.13±5.43b
Sufficient	97	36.6	23.90±2.92c	27.28±3.50	34.80±5.12c
Test and significance			F:8.549	F:1.634	F:5.852
			p:0.000	p:0.197	p:0.003
			c>a=b		c>a=b
Willingness to receive nutritional education					
Yes	150	56.6	23.14±2.94	27.40±3.83	32.76±5.95
No	115	43.4	22.71±3.23	25.94±2.90	34.77±3.49
Test and significance			t:1.106	t:3.397	t:-3.443
			P:0.270	P:0.001	P:0.001
	X±SD	**Min-Max**			
Age (years)	40.02±7.97	23–54			

who had worked for 10–15 years had higher perceived quality of nutritional care mean scores than the nurses who had worked for 5–10 years (Table 1).

A statistically significant difference was determined in the mean scores of attitudes regarding the significance of nutritional evaluation and knowledge of nutritional care and perceived quality of nutritional care according to the participants' status of having received nutritional education. Those who had received nutritional education had higher mean scores ($p<0.05$) (Table 1).

A statistically significant difference was determined in the mean scores related to attitudes toward the significance of nutritional evaluation and perceptions of nutritional care quality according to the perceived level of knowledge of nutritional care. Those who reported their level of knowledge as intermediate had higher mean scores ($p<0.05$). No significant difference was found between participants' perceived level of knowledge of nutritional care mean scores and their level of knowledge of nutritional care ($p>0.05$) (Table 1).

A statistically significant difference was determined in the mean scores of attitudes toward the importance of nutritional evaluation, level of knowledge of nutritional care, and perceived quality of nutritional care according to willingness to receive nutritional education. The participants who wanted to receive nutritional education had higher attitude and knowledge mean scores, while those who did not want to receive education had higher perceived quality of nutritional care mean scores ($p<0.05$) (Table 1).

Total Scores of the Nurses' Nutritional Evaluation, Knowledge, Attitude, and Perception of Care Quality

The mean scores of the participants were found to be 22.95±3.11 for attitudes about the necessity of nutritional evaluation, 26.77±3.53 for level of knowledge of nutritional care, and 33.63±5.12 for perceived quality of nutritional care (Table 2).

Table 2. The mean scores of the study participants related to their attitudes toward the importance of nutritional evaluation, their knowledge of nutritional care, and their per-ception of its quality

	X±SD	Min-Max
Perceptions of the significance of nutritional evaluation	22.95±3.11	17-28
Level of knowledge of nutritional care	26.77±3.53	15-37
Perceived Quality of Nutritional Care	33.63±5.12	21-45

Relationships between the Age of the Nurses and Attitudes about the Significance of Nutritional Evaluation, Level of Knowledge of Nutritional Care, and Nutritional Care Quality Perception

A negative moderate correlation was determined between the age of the nurses and their scores from attitudes regarding the significance of nutritional evaluation and knowledge level related to nutritional care. A low-level correlation was detected between age and the scores from perceived quality of nutritional care. A positive moderate correlation was found between the mean scores of attitudes toward the importance of nutritional evaluation, level of knowledge of nutritional care, and perceived quality of nutritional care. A negative moderate correlation was observed between the level of knowledge of nutritional care and perceived quality of nutritional care (Table 3).

DISCUSSION

The results of this study demonstrated that nurses had a high level of positive attitude toward the importance of nutritional evaluation. In a similar study in the literature, it was reported that nurses' attitudes toward the importance of nutritional evaluation were strong and their perceptions of nutritional knowledge and quality of care were slightly above the average.[1] Similar studies in the literature have also shown that nurses have high positive attitudes toward nutrition.[15,16]

The nurses in the present study were seen to have an average level of knowledge of nutritional care. In a similar study, it was reported that nurses had a moderate level of knowledge

Table 3. Relationships between the age of the nurses and their attitude toward the im-portance of nutritional evaluation, level of knowledge about nutritional care, and per-ceived quality of nutritional care scores

Variables	Age	Attitudes to ward the Importance of Nutritional Evaluation	Level of Knowledge of Nutritional Care	Perceived Quality of Nutritional Care
Age				
r	1			
p				
Attitudes toward the importance of nutritional evaluation				
r	–0.217	1		
p	0.000			
Level of knowledge of nutritional care				
r	–0.260	0.375	1	
p	0.000	0.000		
Perceived quality of nutritional care				
r	0.147	0.323	–0.233	1
p	0.017	0.000	0.000	

of nutritional care.[17] In another study, in which the knowledge levels of nurses regarding enteral and parenteral nutrition practices were evaluated, it was stated that the knowledge of nurses was at a moderate level.[12] Considering the similarity of the current study results to findings in the literature, there can be seen to be a need to increase educational interventions to improve the knowledge level of nurses on nutritional care.

The quality of nutritional care perceived by the nurses in this study was found to be average. In the literature, it has been reported that the care provided to patients in the clinics where they work tends to be evaluated by nurses rather negatively.[1] However, another study found that nurses stated higher evaluation scores for the quality of care in the clinics where they worked.[17] In the light of these results, studies should focus on increasing the perception levels of nurses toward care in order to increase the quality of care.

The current study results showed that nurses with an undergraduate degree had a more positive attitude toward the importance of nutritional evaluation and higher scores for the level of knowledge of nutritional care. In similar studies, it has been found that as the education level of nurses increases, so their knowledge levels and attitudes about nutritional care also increase.[13,18,19] It is thought that the increase in the level of education has a positive effect on the attitudes and knowledge levels of nurses regarding nutritional awareness.

In the present study, it was determined that nurses with 20 years or more of professional experience had higher scores in respect of attitudes toward nutritional assessment, nutritional care knowledge, and perceived quality of care. In a previous study, it was reported that nurses who worked in the unit for 1 year or more had higher knowledge, attitudes, and perceptions of nutritional care.[14] It is thought that the experience gained over a longer working period increases the knowledge, attitudes, and perceptions of nurses and will lead to positive results in the quality of care.

The current study found that nurses working in the palliative care unit, intensive care unit, and chest diseases clinic were seen to have a higher perceived quality of nutritional care. Inadequate nutritional support has been said to be one of the important factors affecting the quality of care in intensive care patients.[8] Previous studies have shown that improving the nutritional care provided by nurses significantly reduces the risk of catheter infection, pressure ulcers, and sepsis in patients.[20] The high level of perception seen in the current study results in units where patients are treated for a long time suggests that this is a factor in improving the quality of care.

In the present study, it was found that the majority of the nurses (53.6%) had previously received nutritional education. From this result, it was determined that the nurses' attitudes regarding the significance of nutritional evaluation, knowledge about nutritional care, and the perceived quality of that care were positively affected by having received nutritional education. Previous studies have reported that the majority of nurses (56.4%, 65.9%, 62%, respectively) had previously received nutritional education, similar to the present study,[1,16,21] and this was thought to have contributed to positive effects on the level of knowledge, attitude, and perception of nutritional care.

The majority of nurses (59.6%) in the current study had a moderate level of nutritional care knowledge, and these nurses had higher attitudes toward the importance of nutritional evaluation and perceived quality of nutritional care. It has been stated that the level of nutritional knowledge of nurses will significantly affect their attitudes and practices regarding the implementation of nutritional care for patients.[2] The fact that the nutritional attitude and quality of care were high in the group representing the majority of nurses in the present study suggests that the quality of nursing care will increase with evidence-based practices in accordance with the literature.

The current study results showed that the majority of nurses (53.6%) had received nutritional education and had more positive attitudes regarding the significance of nutritional assessment and knowledge related to nutritional care and perceived quality of nutritional care. In a previous study, it was stated that the majority of nurses (77.7%) had received nutritional education but this had no effect on attitudes regarding the significance of nutritional evaluation and knowledge related to nutritional care and perceived quality of nutritional care.[14]

A willingness to receive nutritional education was stated by the majority of nurses (56.6%) in this study, and these nurses demonstrated a high level of knowledge regarding nutritional care and reported a high perception of its quality. In another previous study, significance was found between nurses working in units other than intensive care units and the perceived quality of nutritional care.[1] Nurses play an important role in meeting nutritional needs, and this important role depends on their competencies related to nutrition and their level of knowledge about nutritional support.[16] Studies have shown that professional and effective educational practices increase the awareness and competence of nurses toward nutrition, have a favorable effect on their attitudes, and contribute positively to patient care outcomes.[22-24] In the present study, considering the high willingness of the nurses to receive education, it is thought that the education provided will increase the knowledge and perceived quality of nutritional care, resulting in the provision of high-quality care.

CONCLUSION

The results of the present study showed that although nurses have positive attitudes toward the importance of nutritional evaluation, their level of knowledge and perception of quality of nutritional care is moderate. It was also seen that nurses' years of service, working time in the clinic, nutritional education status, nutritional care knowledge level, and willingness to receive nutritional education affect their attitudes regarding the importance attributed to nutritional assessment, their knowledge level in nutritional care, and their perception of its quality. The level of education of the nurses was found to affect their attitudes about the importance of nutritional assessment and their level of knowledge about nutritional care, and the clinics in which the nurses worked affected their perceived quality of nutritional care. In line with these results, it can be recommended that the knowledge of nurses should be questioned and updated at regular intervals, and deficiencies in this regard should be identified and supported with training programs.

Limitations

The main limitation of the study was that it was restricted to nurses working in internal medicine and intensive care units in two district hospitals in a province located in the northern region of Turkey, thereby limiting the generalizability of the findings.

DECLARATIONS

Ethics Committee Approval: The study was approved by Trabzon University Social and Human Sciences Scientific Research and Publication Ethics Committee (No: E-81614018-050.04-2400014629 Date: 18/03/2024).

Informed Consent: Informed consent was obtained from the internal clinic nurses before starting the study.

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: Not declared.

Authorship Contributions: Concept – NIA, SDC; Design – NIA, SDC; Supervision – NIA, SDC; Fundings – NIA, SDC; Materials – NIA, SDC; Data collection &/or processing – NIA, SDC; Analysis and/or interpretation – NIA, SDC; Literature search – NIA, SDC; Writing – NIA, SDC; Critical review – NIA, SDC.

Acknowledgements: The authors would like to thank all the nurses who participated in the study.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Çoşğun T, Kısacık ÖG. Determination the attitude toward the nutritional assessment, the level of knowledge of nutritional care and the perceived quality of care among nurses. Celal Bayar Univ J Inst Health Scı 2021;8:204–17.
2. Kim H, Choue R. Nurses' positive attitudes to nutritional management but limited knowledge of nutritional assessment in Korea. Int Nurs Rev 2009;56:333–9.
3. Shin BC, Chun IA, Ryu SY, Oh JE, Choi PK, Kang HG. Association between indication for therapy by nutrition support team and nutritional status. Medicine (Baltimore) 2018;97:e13932.
4. Barker LA, Gout BS, Crowe TC. Hospital malnutrition: prevalence, identification and impact on patients and the healthcare system. Int J Environ Res Public Health 2011;8:514–27.
5. Gürlek Kısacık Ö, Çoşğun T, Taştekin A. The psychometric properties of the turkish version of the assessment questionnaire of the importance of nutritional assessment, the level of knowledge and perceived quality of nutritional care for nurses. EGE HFD 2019;35:123–35.
6. Philipson TJ, Snider JT, Lakdawalla DN, Stryckman B, Goldman DP. Impact of oral nutritional supplementation on hospital outcomes. Am J Manag Care 2013;19:121–8.
7. Sivrikaya SK, Eryılmaz A. Nutritional supplement in the team nursing. Samsun J Health Sci 2018;3:33–7. [Article in Turkish]
8. Avelino-Silva TJ, Jaluul O. Malnutrition in hospitalized older patients: management strategies to improve patient care and clinical outcomes. Int J Gerontol 2017;11:56–61.
9. Ayık DB, Enç N. Enteral nutrition in intensive care patients. J İntens Care Nurs 2019;23:114–22. [Article in Turkish]
10. Al Kalaldeh M, Shahein M. Nurses' knowledge and responsibility toward nutritional assessment for patients in intensive care units. J Health Sci 2014;4:90–6.
11. Yılmaz D, Sarkut P, Düzgün F, Kuzu C, Kılıçturgay, S. The views of nurses on nutritional assessment and nutritional support of hospitalized patients. J Educ Res Nurs 2017;14:139–43. [Article in Turkish]
12. Koçhan E, Akın S. Assessment of nurses' knowledge levels regarding enteral and parenteral nutrition practices. JAREN 2018;4:1–14.
13. Miriam Theilla RN, Cohen J, Singer P, Liebman C, Kagan I. The assessment, knowledge and perceived quality of nutrition care amongst nurses. J Nutri Med Diet Care 2016;2:012.
14. Aliş B, Okuroğlu GK. Nurses' knowledge, attitudes and perceptions on nutrition in patient care: a descriptive research. Turkiye Klinikleri J Nurs Sci 2025;17:117–25. [Article in Turkish]

15. Bachrach-Lindström M, Jensen S, Lundin R, Christensson L. Attitudes of nursing staff working with older people towards nutritional nursing care. J Clin Nurs 2007;16:2007–14.
16. Karasu M, Özşaker E. An examination of the knowledge, attitudes and practice of nurses regarding the nutrition of surgery patients. Adıyaman Univ J Health Sci 2019;5:1191–205. [Article in Turkish]
17. Yalcin N, Cihan A, Gundogdu H, Ocakci A. Nutrition Knowledge Level of Nurses.Health Science Journal 2013;7(1):99-108.
18. Döngel HE, Kol EÖ, Gönderen K. Attitudes, knowledge, and evaluations of nurses working in training and research hospital regarding nutritional care. Clin Sci Nutrit 2022;4:80–7.
19. Kalender N, Tosun N, Kiliç S. Determination of the level of knowledge of nurses working at an education hospital about total parenteral nutrition. Turkiye Klinikleri J Nurs Sci 2015;7:10–9. [Article in Turkish]
20. Corkins MR, Guenter P, DiMaria-Ghalili RA, Jensen GL, Malone A, Miller S, et al. Malnutrition diagnoses in hospitalized patients: United States, 2010. JPEN J Parenter Enteral Nutr 2014;38:186–95.
21. Kıymaz D, Kılıç Ü, Yücesan S, Öztürk R, Toraman M. The knowledge levels of intensive care nurses on nutritional care. Kırşehir Ahi Evran Univ J Health Sci 2023;7:172–84.
22. Hossaini Alhashemi S, Ghorbani R, Vazin A. Improving knowledge, attitudes, and practice of nurses in medication administration through enteral feeding tubes by clinical pharmacists: a case-control study. Adv Med Educ Pract 2019;10:493–500.
23. Kim H, Chang SJ. Implementing an educational program to improve critical care nurses' enteral nutritional support. Aust Crit Care 2019;32:218–22.
24. Sharour LA. Improving oncology nurses' knowledge, self-confidence, and self-efficacy in nutritional assessment and counseling for patients with cancer: A quasi-experimental design. Nutrition 2019;62:131–4.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2025.32559
Eur Arch Med Res 2026;42(2):133–138

Impact of Single-Event Multilevel Surgery on Energy Expenditure and Gait in Children With Cerebral Palsy

Kadir Cevik,[1] Dilan Deniz Yilmaz,[2] Zeynep Altinkaya,[2] Didem Derici Yildirim,[3] Ugur Dal,[2] Abtullah Milcan[4]

[1]Department of Hand Surgery, Basaksehir Cam and Sakura City Hospital, Istanbul, Turkiye
[2]Department of Physiology, Faculty of Medicine, Mersin University, Mersin, Turkiye
[3]Department of Biostatistics and Medical Informatics, Faculty of Medicine, Mersin University, Mersin, Turkiye
[4]Department of Orthopedics and Traumatology, Mersin University, Faculty of Medicine, Mersin, Turkiye

ABSTRACT

Objective: Pre-pubertal cerebral palsy (CP) patients, particularly those classified as gross motor function classification system (GMFCS) grade II, often develop fixed musculoskeletal contractures that impair gait efficiency and increase energy expenditure. Among available treatment options, single-event multilevel surgery (SEMLS) offers the advantage of addressing multiple deformities in a single operative session, potentially leading to more efficient functional recovery and improved mobility. This study aimed to evaluate the effects of SEMLS on gait characteristics and walking energy expenditure in pre-pubertal children with spastic CP.

Materials and Methods: This study included patients aged 8–13 years who underwent SEMLS between November 2016 and September 2017. All participants had flexion and adduction contractures of the hip, a popliteal angle >50°, and equinus deformity, and were able to walk with or without support. Following long leg plaster cast immobilization for 6 weeks post-surgery, all patients received physical therapy. Three-dimensional gait analysis and walking energy expenditure (oxygen cost) were assessed one month before and 8–12 months after surgery.

Results: Post-operatively, there was a significant reduction in walking energy expenditure, despite no significant change in resting energy expenditure. Although not statistically significant, patients demonstrated an increased preferred walking speed. No significant differences were observed in mediolateral and vertical sacral displacement.

Conclusion: SEMLS significantly reduces walking energy expenditure and improves gait efficiency in pre-pubertal children with spastic CP (GMFCS grade II). These findings support SEMLS as an effective intervention, though larger studies with extended follow-up are needed to confirm long-term outcomes.

Keywords: Cerebral palsy, Children, Energy expenditure, Single-event multilevel surgery, Walking analysis

Cite this article as: Cevik K, Yilmaz DD, Altinkaya Z, Derici Yildirim D, Dal U, Milcan A. Impact of Single-Event Multilevel Surgery on Energy Expenditure and Gait in Children With Cerebral Palsy. Eur Arch Med Res 2026;42(2):133–138.

Address for correspondence: Kadir Cevik. Department of Hand Surgery, Başakşehir Cam and Sakura City Hospital, Istanbul, Türkiye
E-mail: drkadircevik@gmail.com **ORCID ID:** 0000-0003-1606-6728

Submitted: 17.05.2025 **Revised:** 04.08.2025 **Accepted:** 06.08.2025 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

INTRODUCTION

Cerebral palsy (CP) is a group of disorders caused by non-progressive brain injury occurring during early development when brain plasticity is high. CP has several types, including spastic, dyskinetic, ataxic, and mixed forms. In spastic CP, which is the most common, increased muscle tone typically affects the lower limbs (hip flexors, adductors, hamstrings, gastrocnemius), especially in spastic diplegia. According to which part of the brain damaged, whole upper and/or lower extremities can be affected by contractures resulting in a pathological gait pattern and deterioration in gross motor functions.[1] Children with CP often develop fixed contractures in affected muscle groups, leading to pathological gait patterns and reduced functional mobility. Conservative treatments such as physiotherapy, orthoses, and botulinum toxin injections may help manage symptoms but are often insufficient in cases with established deformities. In this context, Single-Event Multilevel Surgery (SEMLS) has become a preferred surgical approach. Unlike traditional orthopedic procedures that target isolated deformities across multiple operations, SEMLS addresses multiple musculoskeletal abnormalities in a single operative session. It aims to optimize gait mechanics, reduce cumulative anesthetic exposure, and shorten rehabilitation time. SEMLS is especially valuable in children with spastic diplegia, where bilateral lower extremity involvement is predominant.[2] Three-dimensional gait analysis (3DGA) serves as a key adjunct in both pre-operative planning and post-operative assessment, offering objective insight into dynamic gait parameters and supporting clinical decision-making.[3,4] Recent evidence further supports the efficacy of SEMLS in improving gait parameters.[5]

The primary aim of this study was to compare the walking energy expenditure and 3DGA of pre-puberty patients with CP who have lower leg extremity contractures and pertaining to level 2 according to Gross Motor Function Classification System (GMFCS), before and after multilevel muscle contracture release, including hip adductor and iliopsoas tenotomy, hamstring release and Achiloplasty. The findings of this study may provide additional insights into the effectiveness of multilevel release operation, by indicating any decrease in walking energy expenditure or improvement in their walking by 3DGA measurements after surgery.

MATERIALS AND METHODS

Participants

Eleven children who can walk on a treadmill without any assistance were selected from 30 CP children applied with SEMLS, and 3 were also excluded later, since they did not attend post-surgery measurements. Then, 8 children aged between 8 and 13 years with a confirmed diagnosis of bilateral spastic CP, involved in level 2 according to GMFCS were evaluated. All participants had been followed by the physical therapy and rehabilitation department for at least 1 year and had received treatments including botulinum toxin injections before being referred for surgical intervention. The study was a self-controlled clinical investigation which included energy expenditure and 3DGA measurements of each participant 1 month before surgery and 8–12 months after surgery. The inclusion criteria were the presence of any flexion and adduction contracture of the hip, a popliteal angle >50°, equinus deformity, ability to walk with/without external support, and oxygen mask tolerability. All subjects and their attendants were well informed about the protocols of the study and written informed consent was obtained. The Declaration of Helsinki was taken into consideration during the study, and the ethical approval of the study was granted from the Local Ethical Committee of Clinical Research (Date: March 23, 2017; Number: 2017/73). This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Surgery

All participants had bilateral adductor longus and iliopsoas tenotomies for hip deformities, z-plasty lengthening of the semitendinosus tendon, fractional lengthening of the semimembranosus and biceps femoris muscles to correct knee flexion deformities, and finally lengthening of the Achilles tendon by the Hoke method for the equinus deformity of the foot.

Surgical procedures were selected based on clinical indications, including the presence of fixed contractures identified through individual gait analysis and physical examination, such as excessive hip adduction, hamstring tightness, and equinus deformity that impaired functional ambulation. Surgical candidates were determined following a multidisciplinary evaluation involving orthopedic surgeons, rehabilitation specialists, and gait analysis teams. An abduction bar above-knee cast was applied for 6 weeks after surgery. Following the cast removal, knee-ankle-foot orthosis was applied and physical therapy initiated.[6]

Energy Expenditure Measurement

Resting and walking energy expenditure were assessed using a validated pediatric indirect calorimetry system, under standardized laboratory conditions (Vmax Spectra 29c. Yorba Linda, CA, USA). All devices were calibrated according to the manufacturer's instructions before each testing session to ensure accuracy. All measurements were done at 08:00–09:30 a.m., with temperature 20–24°C and humidity of 50%.[7] The standing height of subjects was measured by a stadiometer and their body compositions were determined by the bioelectrical impedance analysis method (Tanita BC-418 MA, Tanita Corpo-

ration, Japan). Participants were asked to fast overnight and avoid physical activity during the previous day. Resting energy expenditure (REE) was measured using a pediatric face mask that continuously collected respiratory gas data throughout the assessment. To reduce anxiety and improve cooperation, the presence of a parent and familiar objects was permitted, and the environment was quiet and private. Children were allowed to adapt to the mask for several minutes before beginning the 15-min REE protocol while lying calmly in a supine position. Preferred walking speed (PWS) was determined by having participants perform multiple walking trials on a fixed indoor track. Timing data were collected using motion sensors positioned along the walkway, and the average walking speed was calculated based on these repeated trials. Then, all subjects performed a familiarization protocol to treadmill (Viasys Health Care, USA) for at least 10 min.[7] Subjects rested until their heart rate stabilized within ±5 bpm of baseline.[8] Then, they were instructed to walk for 7 min at their PWS and allowed for holding the supporting bars of the treadmill while they were walking.

The first 10 min of REE measurement was accepted as a habituation period and excluded from analysis following prior pediatric studies using indirect calorimetry.[7] In terms of walking energy expenditure, the past 2 min of energy expenditure record was evaluated as a steady-state period and analyzed as averaging over 10 s intervals. The respiratory exchange ratio was also recorded to evaluate the intensity of exercise.[7]

Walking Analysis Protocol

The 3DGA of subjects was conducted by high-speed CCD cameras (BASLER Vision Technologies, Germany) while they were walking on a treadmill simultaneously with walking energy expenditure. A single marker to the sacrum (on skin), and 2 markers to the mid-points of heels (on shoes) were strapped to subjects for walking analysis. The sacral marker was strapped to the midpoint between two posterior superior iliac spine of patients (Plug-in Gait).

Gait parameters were recorded using a motion capture system, with reflective markers placed on anatomical landmarks. Recorded data were processed using dedicated motion analysis software. The 3DGA system was calibrated before each recording. Step width and stride frequency were evaluated as analyzing the vertical and mediolateral displacement of foot markers. The step width was determined as assuming the minimum level of vertical position of foot markers as heel-contact, then calculating the difference between mediolateral positions of foot markers. The system sensitivity was 0,001 mm.

Statistical Analysis

All data were analyzed using STATISTICA version 13.3 (TIBCO Software Inc., 2017). Descriptive statistics are presented as mean±standard deviation. The Shapiro–Wilk test was employed to assess the normality of the data. For within-subject comparisons between pre- and post-surgery values, the paired samples t-test was used for normally distributed variables, while the Wilcoxon signed-rank test was used for non-normal distributions. Statistical significance was set at $p<0.05$. In addition, effect sizes were calculated to assess the magnitude of observed differences: Cohen's d for parametric tests and rank-biserial correlation (r) for non-parametric tests.

RESULTS

8 CP children (5 boys and 3 girls) with a mean age of 10.88±1.73 years, pertaining to level 2 according to GMFCS, were included in this study. All patients had multilevel contracture release of their lower extremities. One of the patients was excluded from the study as he could not complete the walking analysis protocol. The anthropometric characteristics of subjects before and after surgery are listed in Table 1. Height is the only variable that differed significantly before and after surgery ($p<0.05$) (Table 1). VO_2 (L/min), which is the consumed O_2 in liters per minutes, was not significantly differed during resting condition ($p>0.05$). However; subjects have significantly higher VO_2 (L/min) values during walking before surgery compared to after surgery ($p<0.05$). Another expression for the walking efficiency, O_2 cost (mL/kg/m) which was calculated as normalizing walking O_2 expenditure (L/min) to body weight (kg) of patients, then dividing to PWS (m/min), has significantly decreased after surgery ($p<0.05$). The O_2 cost can provide a straight evaluation of O_2 expenditure per unit of walking distance by eliminating the effect of body weight differences among all participants (Table 2).

Table 1. Anthropometric characteristics of subjects before and after surgery

Variable	Pre-Surgery (mean±SD) n=7	Post-Surgery (8-12 months) (mean±SD) n=7	p
Height (cm)	132.76±11.43	137.63±11.46	<0.001
Body Weight (kg)	28.24±9.92	32.19±13.30	0.051
BMI (kg/m^2)	15.90±2.90	17.17±4.20	0.226
Fat Percent (%)	20.35±5.61	21.88±5.76	0.433
Fat-Free Mass (kg)	22.02±4.42	26.22±8.38	0.112

BMI: Body Mass Index. All values are reported as mean±SD.

Table 2. Comparison of pre-surgery and post-surgery mean±SD values of various parameters

Parameter	Pre-Surgery (mean±SD) n=7	Post-Surgery (8-12 months) (mean±SD) n=7	Effect's size	p
VO_2 during resting (l/min)	0.20±0.05	0.20±0.06	d=0.00	0.811
VO_2 during walking (l/min)	0.53 (0.50-0.78)*	0.51 (0.37-0.72)*	r=0.89	0.012
O_2 cost during walking (ml/kg/m)	1.15±0.52	0.72±0.31	d=1.04	0.011
PWS (km/h)	1.39±0.49	1.56±0.45	d=0.36	0.093
Cadence (steps/min)	63.29±22.68	65.00±17.66	d=0.08	0.630
Step width (mm)	119.84±23.78	136.55±56.58	d=0.42	0.386
Displacement of sacral marker in mediolateral direction (mm)	26.84±7.99	22.38±6.07	d=0.62	0.157
Displacement of sacral marker in vertical direction (mm)	24.51±8.29	25.84±6.46	d=0.18	0.737

PWS: preferred walking speeds; VO_2: O_2 volume ; Values are reported as mean±SD except * indicates median (Q1-Q3).

In terms of gait variables of subjects, PWS (km/h) was not changed significantly after surgery compared to before surgery ($p>0.05$). Similarly, cadence (steps/minute) and step width (mm) were not increased significantly after surgery ($p>0.05$). The pre- and post-surgery differences in mediolateral and vertical displacement of the sacral marker were also not changed significantly after surgery ($p>0.05$) (Table 2).

Post-operative improvements were statistically significant and clinically meaningful for O_2 cost during walking (Cohen's $d=1.04$) and VO_2 during walking (rank-biserial $r=0.89$), both indicating large effect sizes.

Although changes in gait parameters such as PWS, cadence, step width, and sacral displacement were not statistically significant, these findings may reflect individual variability or limited post-operative observation duration.

DISCUSSION

This study evaluates energy expenditure while walking with 3DGA before and after multilevel release surgery in children with CP to evaluate its efficiency. Walking intensity significantly affects energy consumption, since quicker walking requires higher muscle performance and energy expenditure.[9] In their investigation, Rose et al.[10] discovered that CP patients used 2–3 times more energy walking in PWS than healthy volunteers. Our participants' mean walking energy consumption dropped considerably after SEMLS. Dahlbäck et al.[11] showed a 5% decreased oxygen demand per unit body weight after surgery. Another study of 135 CP patients found that O_2 consumption reduced 9% in orthopedic surgery (only) and 25% in orthopedic surgery with rhizotomy. The "Functional Assessment Query" showed a 38% improvement in gait pathology in the group who had both procedures.[12]

Increased body mass can raise the metabolic cost of walking, which may affect oxygen consumption. O_2 cost, which accounts for body weight differences, showed significant improvement post-operatively, consistent with prior literature.[9]

We observed that gait analysis improved our patients' mean PWS following surgery, although not significantly. Despite surgery, key gait metrics such as cadence and step width showed no significant short-term changes, aligning with earlier reports. We found that these individuals consumed less energy throughout walking tests without changing gait characteristics. However, after surgery, individuals prefer to walk faster with increased stride frequency and step width, reduced mediolateral sacral marker oscillation, and increased vertical oscillation. Individual adaptation time variation, a limited sample size, and inadequate statistical power may explain these factors' lack of statistical significance. Others may have modest dynamic gait alterations using knee-ankle-foot orthoses.

Only three markers were used in the gait investigation, limiting the biomechanical interpretation of hip, knee, and ankle joint kinematics. For functional gait evaluation, cadence and step width were prioritized. This technique met the study's main goal; however, future research should include comprehensive 3D gait analysis and reliability checks for broader interpretation.

Literature has reported mixed results on SEMLS effects on gait parameters in children with CP. Factors influencing short-term recovery after SEMLS have also been identified in recent studies, further highlighting the variability in patient outcomes.[13] Cadence improved in 18%[14-16] and walking speed in 24% of cases.[14,17-20] Dreher et al.[21] observed significant improvements in scores worsened at 1 year. Steinwender et al.[17] noted

a gradual cadence decrease but improved walking speed. Kay et al.[22] and Graham et al.[23] reported minimal or no changes, particularly in older children. Similarly, Bernthal et al.[24] observed modest gains in walking speed and cadence following soft-tissue surgery, with unchanged pelvic kinematics but improved hip extension. Harvey et al.[25] documented mobility returning to baseline by 12 months, with improvements evident after 2 years. Cruz et al.[26] and Presedo et al.[27] reported no significant changes in walking speed after distal rectus femoris procedures, though knee kinematics improved.

Consistent with this, our study found no significant short-term change in sacral marker displacement. Rutz et al.[28] similarly reported no major gait parameter changes within 1–3 years of SEMLS, but notable improvements at 4–6 years. Svehlík et al.[29] and Dreher et al.[21] have emphasized the importance of mid- to long-term follow-up to accurately capture functional gains. These findings suggest that short-term assessments may underestimate the long-term benefits of SEMLS.

Our sacral marker results are consistent with previous studies showing minimal short-term changes in pelvic and hip joint kinematics after SEMLS.[16] McMulkin et al.[30] further reported that combining soft tissue and bony procedures can lead to greater improvements in gait metrics, underscoring the variability of outcomes depending on surgical approach and follow-up duration.

CONCLUSION

Taken together, the findings of this study contribute to the growing body of evidence supporting the metabolic benefits of SEMLS in children with spastic CP. While many previous studies have focused on long-term functional or kinematic improvements, our study uniquely highlights a significant reduction in walking energy expenditure shortly after surgery, despite the absence of statistically significant changes in gait parameters such as PWS, cadence, and step width. These results suggest that improved energy efficiency may serve as an early and sensitive indicator of surgical benefit and underscore the importance of incorporating both gait analysis and metabolic assessments in post-operative evaluation.

This study has several limitations that should be considered when interpreting the findings. First, the sample size was small, as only a limited number of patients were eligible and tolerated the procedures, particularly the energy expenditure testing using a face mask. Second, the follow-up period was relatively short, which may have precluded the detection of long-term functional or kinematic improvements. Third, the gait analysis used only three reflective markers and a limited number of cameras, which may have restricted the accuracy and comprehensiveness of joint kinematic data.

DECLARATIONS

Ethics Committee Approval: The study was approved by Mersin University Ethics Committee (No: 2017/73 Date: 23/03/2017).

Informed Consent: Written informed consent was obtained.

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: We disclose that we don't use artificial intelligence (AI).

Authorship Contributions: Concept – KC; Design – KC; Supervision – AM; Materials – DDY; Data collection &/or processing – ZA; Analysis and/or interpretation – DDYıldırım; Literature search – KC; Writing – KC; Critical review – UD.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Rosenbaum P, Paneth N, Leviton A, Goldstein M, Bax M, Damiano D, et al. A report: the definition and classification of cerebral palsy April 2006. Dev Med Child Neurol Suppl 2007;109:8–14. Erratum in: Dev Med Child Neurol 2007;49:480.
2. McGinley JL, Dobson F, Ganeshalingam R, Shore BJ, Rutz E, Graham HK. Single-event multilevel surgery for children with cerebral palsy: a systematic review. Dev Med Child Neurol 2012;54:117–28.
3. Cook RE, Schneider I, Hazlewood ME, Hillman SJ, Robb JE. Gait analysis alters decision-making in cerebral palsy. J Pediatr Orthop 2003;23:292–5.
4. Mueske NM, Õunpuu S, Ryan DD, Healy BS, Thomson J, Choi P, et al. Impact of gait analysis on pathology identification and surgical recommendations in children with spina bifida. Gait Posture 2019;67:128–32.
5. Moisan G, Bonnefoy-Mazure A, De Coulon G, Tabard-Fougère A, Armand S, Turcot K. Assessment of gait quality and efficiency after undergoing a single-event multilevel surgery in children with cerebral palsy presenting an intoeing gait pattern. Childs Nerv Syst 2022;38:1523–30.
6. Dal U, Erdogan T, Resitoglu B, Beydagi H. Determination of preferred walking speed on treadmill may lead to high oxygen cost on treadmill walking. Gait Posture 2010;31:366–9.
7. Compher C, Frankenfield D, Keim N, Roth-Yousey L; Evidence Analysis Working Group. Best practice methods to apply to measurement of resting metabolic rate in adults: a systematic review. J Am Diet Assoc 2006;106:881–903.
8. Van de Putte M, Hagemeister N, St-Onge N, Parent G, de Guise JA. Habituation to treadmill walking. Biomed Mater Eng 2006;16:43–52.

9. Stoquart G, Detrembleur C, Lejeune T. Effect of speed on kinematic, kinetic, electromyographic and energetic reference values during treadmill walking. Neurophysiol Clin 2008;38:105–16.
10. Rose J, Gamble JG, Burgos A, Medeiros J, Haskell WL. Energy expenditure index of walking for normal children and for children with cerebral palsy. Dev Med Child Neurol 1990;32:333–40.
11. Dahlbäck GO, Norlin R. The effect of corrective surgery on energy expenditure during ambulation in children with cerebral palsy. Eur J Appl Physiol Occup Physiol 1985;54:67–70.
12. Saraph V, Zwick EB, Zwick G, Steinwender C, Steinwender G, Linhart W. Multilevel surgery in spastic diplegia: evaluation by physical examination and gait analysis in 25 children. J Pediatr Orthop 2002;22:150–7.
13. Lennon N, Gerry G, Biermann I, Beaman J, Mamula N, Gilmore A, et al. Factors associated with short-term recovery following single-event multilevel surgery for children with cerebral palsy. Pediatr Phys Ther 2023;35:93–9.
14. Saraph V, Zwick EB, Auner C, Schneider F, Steinwender G, Linhart W. Gait improvement surgery in diplegic children: how long do the improvements last? J Pediatr Orthop 2005;25:263–7.
15. Adolfsen SE, Ounpuu S, Bell KJ, DeLuca PA. Kinematic and kinetic outcomes after identical multilevel soft tissue surgery in children with cerebral palsy. J Pediatr Orthop 2007;27:658–67.
16. Zwick EB, Saraph V, Linhart WE, Steinwender G. Propulsive function during gait in diplegic children: evaluation after surgery for gait improvement. J Pediatr Orthop B 2001;10:226–33.
17. Steinwender G, Saraph V, Zwick EB, Uitz C, Linhart W. Assessment of hip rotation after gait improvement surgery in cerebral palsy. Acta Orthop Belg 2000;66:259–64.
18. Dreher T, Buccoliero T, Wolf SI, Heitzmann D, Gantz S, Braatz F, et al. Long-term results after gastrocnemius-soleus intramuscular aponeurotic recession as a part of multilevel surgery in spastic diplegic cerebral palsy. J Bone Joint Surg Am 2012;94:627–37.
19. Dreher T, Vegvari D, Wolf SI, Geisbüsch A, Gantz S, Wenz W, et al. Development of knee function after hamstring lengthening as a part of multilevel surgery in children with spastic diplegia: a long-term outcome study. J Bone Joint Surg Am 2012;94:121–30.
20. de Freitas Guardini KM, Kawamura CM, Lopes JAF, Fujino MH, Blumetti FC, de Morais Filho MC. Factors related to better outcomes after single-event multilevel surgery (SEMLS) in patients with cerebral palsy. Gait Posture 2021;86:260–5.
21. Dreher T, Thomason P, Švehlík M, Döderlein L, Wolf SI, Putz C, et al. Long-term development of gait after multilevel surgery in children with cerebral palsy: a multicentre cohort study. Dev Med Child Neurol 2018;60:88–93.
22. Kay RM, Rethlefsen SA, Dennis SW, Skaggs DL. Prediction of postoperative gait velocity in cerebral palsy. J Pediatr Orthop B 2001;10:275–8.
23. Graham HK, Baker R, Dobson F, Morris ME. Multilevel orthopaedic surgery in group IV spastic hemiplegia. J Bone Joint Surg Br 2005;87:548–55.
24. Bernthal NM, Gamradt SC, Kay RM, Wren TA, Cuomo AV, Reid J, et al. Static and dynamic gait parameters before and after multilevel soft tissue surgery in ambulating children with cerebral palsy. J Pediatr Orthop 2010;30:174–9.
25. Harvey A, Graham HK, Morris ME, Baker R, Wolfe R. The Functional Mobility Scale: ability to detect change following single event multilevel surgery. Dev Med Child Neurol 2007;49:603–7.
26. Cruz AI, Ounpuu S, Deluca PA. Distal rectus femoris intramuscular lengthening for the correction of stiff-knee gait in children with cerebral palsy. J Pediatr Orthop 2011;31:541–7.
27. Presedo A, Megrot F, Ilharreborde B, Mazda K, Penneçot GF. Rectus femoris distal tendon resection improves knee motion in patients with spastic diplegia. Clin Orthop Relat Res 2012;470:1312–9.
28. Rutz E, Baker R, Tirosh O, Brunner R. Are results after single-event multilevel surgery in cerebral palsy durable? Clin Orthop Relat Res 2013;471:1028–38.
29. Svehlík M, Steinwender G, Kraus T, Saraph V, Lehmann T, Linhart WE, et al. The influence of age at single-event multilevel surgery on outcome in children with cerebral palsy who walk with flexed knee gait. Dev Med Child Neurol 2011;53:730–5.
30. McMulkin ML, Gordon AB, Caskey PM, Tompkins BJ, Baird GO. Outcomes of orthopaedic surgery with and without an external femoral derotational osteotomy in children with cerebral palsy. J Pediatr Orthop 2016;36:382–6.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2025.08216
Eur Arch Med Res 2026;42(2):139–147

Radiological BI-RADS Scoring and Histopathological Results: Correlation and Pitfalls

Selcuk Cin,[1] Gul Gizem Pamuk,[2] Muhammed Edip İsleyen,[3] Ozgecan Gundogar,[4] Eda Cingoz,[2] Murat Yuce,[5] Merve Cin,[6] Ali Muhammedoglu,[7] Ahmet Tan Cimilli[2]

[1]Department of Pathology, Cerrahpasa Faculty of Medicine, Istanbul University-Cerrahpasa, Istanbul, Turkiye
[2]Department of Radiology, University of Health Sciences, Bağcılar Training and Research Hospital, Istanbul, Turkiye
[3]Department of Pathology, Ceylanpinar State Hospital, Sanlıurfa, Türkiye
[4]Department of Pathology, University of Health Sciences, Gaziosmanpasa Training and Research Hospital, Istanbul, Turkiye
[5]Biomedical Engineering and Imaging Institute, Icahn School of Medicine at Mount Sinai, New York, USA
[6]Department of Pathology, University of Health Sciences, Istanbul Training and Research Hospital, Istanbul, Turkiye
[7]Department of Pathology, University of Health Sciences, Bagcilar Training and Research Hospital, Istanbul, Turkiye

ABSTRACT

Objective: The Breast Imaging Reporting and Data System (BI-RADS) is an internationally utilized radiological reporting system for the classification of breast lesions, categorizing cases based on increasing malignancy rates. Given that breast cancer is the most common type of cancer among women and that early diagnosis and treatment are of paramount importance, the correlation between histomorphological changes and the BI-RADS scoring system holds critical significance. In this study, we aimed to evaluate the histopathological correlation with the BI-RADS scoring system and identify potential diagnostic pitfalls.

Materials and Methods: The radiological findings from breast imaging of 881 women over the age of 40 were retrospectively reviewed and reclassified according to the updated BI-RADS scoring system. The classifications were then correlated with the histopathological results of the biopsies.

Results: In our study, the positive predictive values for detecting malignancy increased progressively from BI-RADS category 4A to 4C, with rates of 6.2%, 34.3%, and 85.7%, respectively. This rate further increased to 94% in category 5.

Conclusion: Although the BI-RADS scoring system, which is effectively used in breast biopsies today, demonstrates a high radiological and pathological correlation, new and innovative studies focusing on subtypes within category 3 and category 4 will contribute to reducing discrepancies observed in these categories.

Keywords: BI-RADS, Breast, Pathology, Radiology

Cite this article as: Cin S, Pamuk GG, Isleyen ME, Gundogar O, Cingoz E, Yuce M, et al. Radiological BI-RADS Scoring and Histopathological Results: Correlation and Pitfalls. Eur Arch Med Res 2026;42(2):139–147.

Address for correspondence: Selcuk Cin. Department of Pathology, Cerrahpaşa Faculty of Medicine, İstanbul University-Cerrahpaşa, İstanbul, Türkiye

E-mail: selc2049@hotmail.com **ORCID ID:** 0000-0001-5097-0505

Submitted: 04.08.2025 **Revised:** 23.08.2025 **Accepted:** 10.09.2025 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

INTRODUCTION

The Breast Imaging Reporting and Data System (BI-RADS), established by the American College of Radiology (ACR) in the late 1980s, aimed to standardize reporting in breast imaging and to improve communication and research in the field. It has become a standardized system for reporting breast pathology detected by mammography, ultrasound, and magnetic resonance imaging. This comprehensive system provides a structured approach to reporting, ensuring consistency in the interpretation of imaging results. At the heart of BI-RADS is a detailed lexicon, a dictionary-like collection of terms that describe specific imaging features. This lexicon is based on data and research to effectively differentiate between benign and malignant breast conditions, thereby enhancing the accuracy and reliability of breast imaging assessments. Since its first edition in 1993, BI-RADS has undergone several revisions, with subsequent editions released in 1995, 1998, and 2003, each adding important elements for clarification, management, and quality assurance.[1] The fourth edition in 2003 introduced a significant enhancement to the BI-RADS assessment category 4.[2] It subdivided this category into three levels: 4A, indicating low suspicion for malignancy; 4B, indicating intermediate suspicion; and 4C, indicating moderate concern but not definitively malignant. This refinement aimed to improve risk stratification, offering more precise assessment and reporting to aid both physicians and patients in understanding probable biopsy results and subsequent follow-up steps. The fifth edition of BI-RADS, released in 2013, brought about significant updates.[3] It eliminated underutilized or redundant findings from the lexicon, thereby streamlining the system. Additionally, it incorporated newer technologies such as shear-wave elastography, making it more contemporary and user-friendly.[4] This latest revision reflects ongoing efforts to enhance the system's relevance and ease of use in breast imaging practice.

The structured lexicon of BI-RADS enhances interaction among various medical professionals involved in complex clinical care. The BI-RADS final assessment categories, coupled with management recommendations, have set a standard in breast care decision-making based on imaging. This uniformity is particularly beneficial in multidisciplinary breast care teams, including surgeons, pathologists, oncologists, and radiologists. For instance, the specific BI-RADS categories such as 4A, 4B, 4C, and 5 serve as valuable tools for conveying the degree of suspicion to pathologists for lesions undergoing imaging-guided biopsies. This promotes accurate histologic diagnoses by fostering a correlation between imaging and pathology. Moreover, clear communication between pathologists and radiologists is crucial for identifying potential sampling errors in percutaneous core biopsies and preventing delayed cancer detection by facilitating timely excisional biopsies.

MATERIALS AND METHODS

Definitions of the Study Groups and Clinical/Pathologic/Radiologic Parameters

Women aged 40 years and older who underwent tru-cut biopsy for breast lesions in the radiology department and received a diagnosis in the pathology clinic at Istanbul Bağcılar Training and Research Hospital between 2020 and 2022 were retrospectively evaluated. The lesions were evaluated using ultrasound (US), mammography, and magnetic resonance imaging (MRI) available in the hospital system and were categorized as BI-RADS 2, 3, 4A, 4B, 4C, or 5. A total of 881 cases were included in the study. Cases with non-diagnostic image quality (mammography, MRI, US) and those that underwent only US examination but had reports that did not conform to the BI-RADS categorization were excluded from the study. Radiologically described lesions were measured in millimeters based on their largest dimension and were classified according to their location as the upper outer quadrant, upper inner quadrant, lower outer quadrant, or lower inner quadrant of the right or left breast. Lesions that did not fully correspond to a specific quadrant, including those located at the 3, 6, 9, and 12 o'clock positions, as well as those in the retroareolar region, were categorized separately. Clinical information of the cases was obtained from the hospital automation system. Histopathological evaluations were classified into six groups based on the morphological findings of the tru-cut biopsies: malignant tumors (regardless of subtype), pure in situ carcinomas (lobular/ductal), fibroadenomas/benign phyllodes tumors (benign fibroepithelial tumors), papillary neoplasms, lipomatous tumors, and fibrocystic changes/benign breast tissue (FCC/BBT), including adenosis, inflammatory conditions, and other non-neoplastic alterations.

Statistical Analysis

A retrospective analysis was conducted on a dataset comprising 881 breast lesions. Descriptive statistics summarized patient demographics, lesion characteristics, pathological outcomes, and radiological classifications based on the BI-RADS categories. Continuous variables, such as patient age and lesion diameter, are presented as means±standard deviations, while categorical variables are expressed as frequencies and percentages. To evaluate differences in mean ages and tumor diameters across different pathological diagnoses, the Kruskal–Wallis test was employed due to the non-parametric distribution of the data. Post hoc analyses identified specific group differences, with significance levels set at $p<0.05$. Chi-square tests assessed associations between pathological diagnoses and both BI-RADS categories and lesion locations within breast quadrants. Adjusted residuals identified significant deviations from expected frequencies, highlighting

specific associations between pathology types and radiological classifications or anatomical locations.

Ethical Approval

This study was conducted in accordance with the principles outlined in the Declaration of Helsinki and received approval from the Non-Interventional Clinical Research Ethics Committee of Bağcılar Training and Research Hospital (Decision No:2023/08/08/043, dated August 8, 2023).

RESULTS

The study evaluated 881 breast lesions with a mean patient age of 51.3±9.9 years. The average lesion diameter was 18.5±12.3 mm. Lesion laterality was nearly balanced, with 469 lesions (53.2%) located in the left breast and 412 lesions (46.8%) in the right breast.

The distribution of pathological diagnoses revealed that fibroadenoma/benign phyllodes tumors were the most common lesion type, identified in 317 cases (36%). Invasive carcinoma was present in 282 lesions (32%), while FCC/BBT was found in 238 lesions (27%). Less frequent pathologies included papillary neoplasms in 24 lesions (2.7%), lipomatous tumors in 13 lesions (1.5%), and pure ductal carcinoma in situ (DCIS) in 7 lesions (0.8%).

Radiologically, the lesions were classified based on BI-RADS categories as follows: BI-RADS 3 in 316 lesions (35.9%), BI-RADS 5 in 244 lesions (27.7%), BI-RADS 4A in 145 lesions (16.5%), BI-RADS 2 in 99 lesions (11.2%), BI-RADS 4C in 42 lesions (4.8%), and BI-RADS 4B in 35 lesions (4%). The distribution of the cases according to BI-RADS categories and the relationships between BI-RADS categories, age, and lesion size are shown in Table 1.

The Kruskal–Wallis test indicated statistically significant differences in mean ages across categories of pathological diagnosis ($p<0.001$). Post hoc analysis revealed that patients with invasive ductal carcinoma had a significantly higher mean age (55.9 years) compared to those with fibroadenoma (47.6 years) and FCC/BBT (50.4 years) ($p<0.001$).

Significant differences were also observed in tumor diameters across pathological diagnoses ($p<0.001$). Post hoc analysis showed that patients with invasive carcinomas had significantly larger mean tumor diameters (24.2 mm) compared to those with fibroadenoma (15.7 mm), papillary neoplasms (16.2 mm), and in situ carcinomas (9.4 mm) ($p<0.001$).

An analysis of lesion distribution across breast quadrants revealed significant associations between pathological diagnoses and specific anatomical locations ($p<0.05$). At the 9 o'clock position, FCC/BBT was significantly more frequent, representing 11.8% of all FCC/BBT cases, while invasive carcinomas were significantly less frequent at this location (3.2% of invasive cases). Cases with diffuse involvement (described as more than one quadrant) exhibited a higher frequency of FCC/BBT (19.3% of these cases) and a lower frequency of fibroadenomas (6.3% of fibroadenoma cases).

In the upper inner quadrant (UIQ), invasive carcinomas were significantly more common, accounting for 16.7% of invasive cases, whereas FCC/BBT was less frequent in this quadrant (8% of cases). Similarly, in the upper outer quadrant (UOQ), invasive carcinomas had the highest frequency, representing 38.7% of invasive cases, while FCC/BBT lesions were less common in this region (24.4% of FCC/BBT cases) (Fig. 1).

The chi-square test revealed a significant association between pathological diagnoses and BI-RADS categories ($p<0.001$). FCC/BBT cases were significantly more frequent in BI-RADS 2, comprising 69.7% of lesions in this category. Fibroadenoma/benign phyllodes tumors were predominantly classified as BI-RADS 3 and BI-RADS 4A, accounting for 66.8% and 25.6% of lesions in these categories, respectively. Invasive carcinomas were significantly associated with higher BI-RADS categories, particularly BI-RADS 5 and BI-RADS 4C (93.7% of all cases). Papillary neoplasms showed a significant association with BI-RADS 4B, comprising 20.0% of lesions in this category.

The malignant cases consisted of invasive carcinomas, pure in situ carcinomas, and papillary lesions, including invasive papillary carcinoma, encapsulated papillary carcinoma, and papilloma containing ductal carcinoma in situ. In our study, there were 145 cases evaluated as BI-RADS 4A, of which 9 were malignant

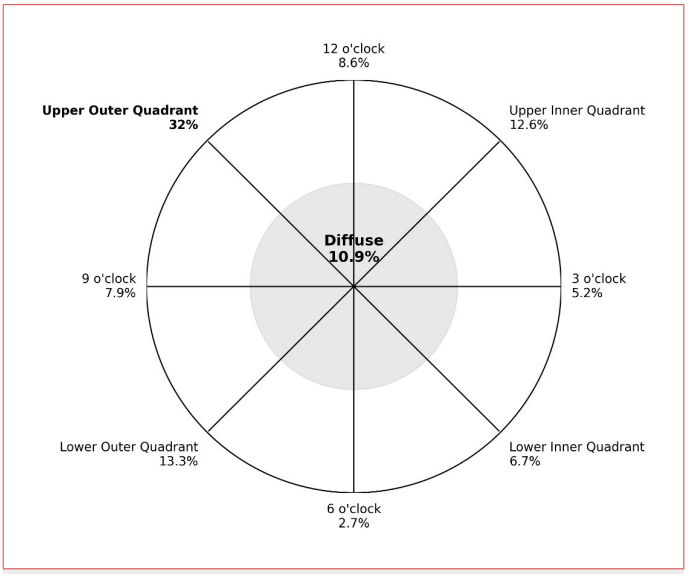

Figure 1. Distribution of all cases by quadrants (*'diffuse' described as more than one quadrant*).

Table 1. Distribution of BI-RADS classification, patient age, and lesion diameter according to pathological diagnosis.[41,48]

Pathology	Median Age (Min, Max)	BI-RADS Class	n	%	Lesion Diameter (Mean±Std)
Fibrocystic changes/benign breast tissue (n=238)	49 (40.78)	BIRADS 2	69	29.0	17.51±1.45
		BIRADS 3	90	37.8	13.24±0.66
		BIRADS 4A	52	21.8	18.03±1.43
		BIRADS 4B	15	6.3	12.87±1.92
		BIRADS 4C	4	1.7	15.4±4.09
		BIRADS 5	8	3.4	18.75±2.87
Invasive carcinomas (n=282)	54 (40.94)	BIRADS 3	3	1.1	15.33±2.91
		BIRADS 4A	7	2.5	20.71±2.52
		BIRADS 4B	8	2.8	23.71±6.93
		BIRADS 4C	34	12.1	22.18±3.1
		BIRADS 5	230	81.6	24.72±0.8
Fibroadenoma/benign filloides tumors (n=317)	45 (40.79)	BIRADS 2	15	4.7	23.62±7.04
		BIRADS 3	211	66.6	14.07±0.5
		BIRADS 4A	81	25.6	17.25±0.88
		BIRADS 4B	5	1.6	12.6±2.89
		BIRADS 4C	2	0.6	65±35
		BIRADS 5	3	0.9	19.5±8.32
Papillary Neoplasms (n=24)	49 (40.94)	BIRADS 2	1	4.2	8
		BIRADS 3	8	33.3	8.31±2.58
		BIRADS 4A	4	16.7	8.88±1.16
		BIRADS 4B	7	29.2	29.14±8.62
		BIRADS 4C	2	8.3	25.5±10.5
		BIRADS 5	2	8.3	11.5±3.5
In situ carcinomas (n=7)	52 (41.71)	BIRADS 2	1	14.3	7
		BIRADS 3	4	57.1	10.2±2.38
		BIRADS 4A	1	14.3	6
		BIRADS 5	1	14.3	12
Lipomatous Tumors (n=13)	53 (40.64)	BIRADS 2	13	100.0	24.24±6.78

(6.21%). The number of BI-RADS 4B cases was 35, with 12 being malignant (34.29%). For cases classified as BI-RADS 4C, there were 42 in total, of which 36 (85.71%) were malignant. Additionally, there were 95 cases categorized as BI-RADS 5, and 90 of these were pathologically confirmed as malignant (94.74%).

The chi-square analysis demonstrated significant associations between specific pathological diagnoses and BI-RADS categories. Patients diagnosed with invasive carcinoma tended to be older and have larger tumors, and their lesions were more likely to be classified as BI-RADS 4C and BI-RADS 5 and located in the upper quadrants of the breast. Fibroadenomas were predominantly associated with BI-RADS 3 and BI-RADS 4A and did not show a strong predilection for specific breast quadrants, while FCC/BBT lesions were more commonly classified as BI-RADS 2 and had higher frequencies at certain anatomical locations, such as the 9 o'clock position (11.8% of all FCC/BBT cases) and in diffuse presentations.

DISCUSSION

Breast carcinomas represent the most frequently diagnosed type of cancer in women.[5] Alongside increasing awareness

among individuals, the widespread adoption of diverse imaging modalities has enabled the early detection of lesions, resulting in a rise in both the number and variety of biopsies obtained from these patients. The findings derived from imaging performed for this purpose are categorized into risk groups using the BI-RADS (Breast Imaging Reporting and Data System) classification, which is employed worldwide and serves as one of the cornerstones in patient management.

Breast core biopsies, which hold a significant place in routine pathology practice, play a crucial role in both diagnosis and treatment, making them an indispensable component of contemporary medical practice. In the histopathological evaluation of breast core biopsies, the integration of patients' clinical information and findings with radiological data is an essential requirement for an accurate diagnosis. In this context, pathologists correlate the microscopic changes observed with the BI-RADS category indicated in imaging studies. Apart from category 5, which is reported to have a malignancy likelihood exceeding 95%, malignant cases can also be encountered in category 4 and, to a lesser extent, in category 3. Particularly in category 4, which is subdivided into 4A, 4B, and 4C with progressively increasing malignancy potential, cases that do not exhibit overt malignancy histopathologically are a cause for concern among pathologists due to the risk of missing a malignant lesion. In this context, there has been a noticeable increase in studies focusing specifically on the BI-RADS 4 group in recent years.[6–8]

In our study, 9 of 145 cases classified as BI-RADS 4A were reported as malignant. Among these, 7 were diagnosed as invasive carcinoma, 1 as ductal carcinoma in situ, and 1 as a papilloma containing ductal carcinoma in situ. Pathological evaluation of 35 cases classified as BI-RADS 4B revealed 8 cases of invasive carcinoma and 7 cases of papillary proliferation. Among these papillary lesions, 4 were primarily evaluated as malignant (invasive papillary carcinoma, encapsulated papillary carcinoma, and papilloma containing ductal carcinoma in situ), while 3 were assessed as intraductal papillomas. Of the 42 cases classified as BI-RADS 4C, 34 were identified as invasive carcinoma, while 2 were evaluated as consistent with encapsulated papillary carcinoma. It was observed that the positive predictive value (PPV) for malignancy in BI-RADS category 4 increased across subcategories. The increasing PPV rates from 4A to 4C were 6.2%, 34.3%, and 85.7%, respectively. The findings reported by Yoon et al.[9] indicated rates of 7.6%, 37.8%, and 81.9%, respectively. In another study, the rates were found to be 13.6%, 50%, and 86%.[10] Kim et al.[11] conducted an analysis in which the PPV was calculated as 9.4% for category 4A and 21.4% for category 4B. Similarly, Bent et al.[12] reported PPV values of 13% for category 4A, 36% for category 4B, and 79% for category 4C. The results we obtained are consistent with the literature, with this rate reaching 94% in category 5. However, considering that the final objective is 100%, the potential reasons underlying this discrepancy require further investigation. Even though these rates and the "discordance" may appear minor, they are quite significant and concerning in daily clinical practice for a tertiary hospital.

While BI-RADS 4C and 5 lesions are considered highly suggestive of malignancy, a portion of these lesions are ultimately benign, highlighting the challenges of imaging-based classification. This discrepancy arises from multiple factors, including imaging characteristics, clinical context, and differences in interpretation. Suspicious ultrasound features such as spiculated margins, microcalcifications, nonparallel orientation, microlobulated contours, posterior shadowing, and complex echogenicity are often associated with malignancy; however, their presence does not always indicate a malignant process.[13] Additionally, patient-related factors such as age, prior cancer history, multiple lesions, and clinical symptoms can influence how lesions are categorized.[9,14] Another contributing factor is variability in radiological interpretation, as differences in experience can lead to inconsistencies. While malignancy rates are expected to exceed 50% in BI-RADS 4C and 95% in BI-RADS 5 lesions, a subset of benign lesions still falls into these categories.[15] All these considerations indicate that current breast imaging techniques and the BI-RADS classification system are still insufficient for achieving optimal diagnostic predictability, as radiology alone cannot always provide a definitive diagnosis, with some BI-RADS 4C and 5 lesions still proving to be benign. In this study, among the cases classified as BI-RADS 4C but not diagnosed with invasive carcinoma, 2 were evaluated as encapsulated papillary carcinoma, 2 as fibroadenoma with cellular stroma/benign phyllodes tumor (Fig. 2), 1 as granulomatous mastitis, 1 as fat necrosis, 1 as chronic nonspecific inflammation, and 1 as microcalcification foci.

Among the 13 cases classified as BI-RADS 5 but not diagnosed with invasive carcinoma or pure in situ carcinoma, there were 2 intraductal papillomas, 1 case of granulomatous mastitis (Fig. 3), 2 cases of fat necrosis, and 3 fibroadenomas. In this context, although the number of discordant cases in the 4C and 5 groups is limited, it can be suggested that cellular fibroadenomas, signs of inflammation, and areas of fat necrosis are the primary contributors to this discordance. Certainly, atypical proliferations may be observed within or at the periphery of a fibroadenoma or an inflammatory lesion, and these atypical areas may not be present in the core biopsy sample. However, in cases of potential discordance,

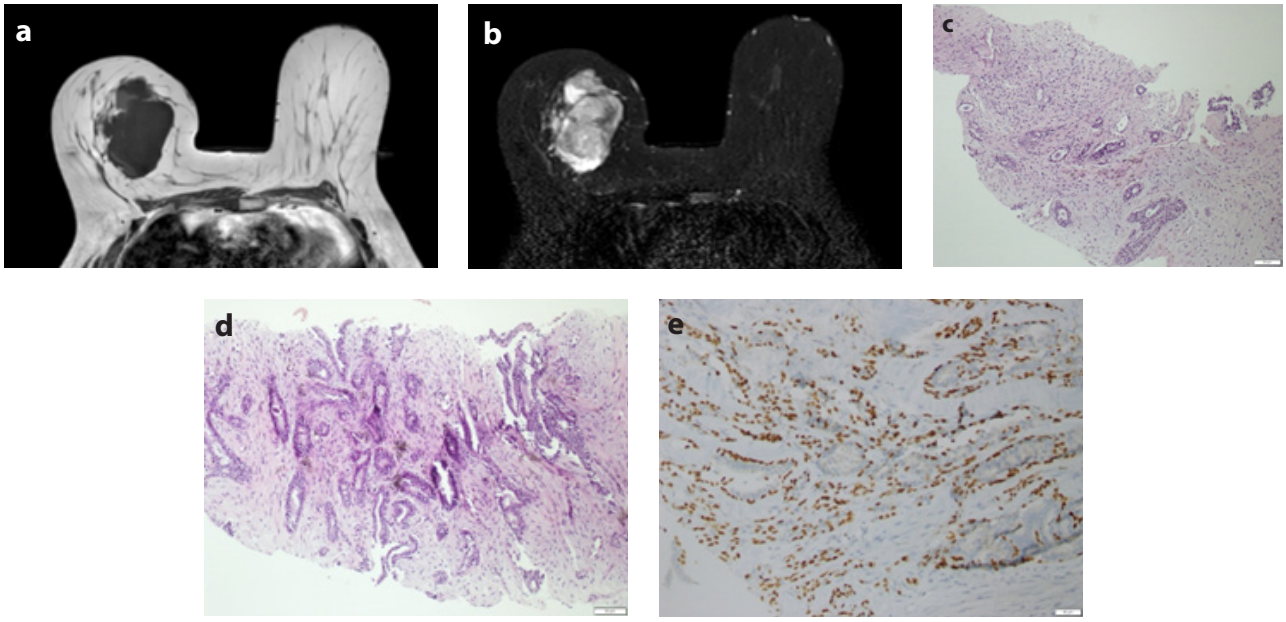

Figure 2. The lesion in the middle-inner section of the right breast, which has an irregular medial contour and is radiologically classified as BI-RADS 4C in conjunction with ultrasound, is shown in T1-weighted **(a)** and post-contrast fat-suppressed **(b)** MRI images. H&E sections reveal glands with a benign appearance, embedded in a mildly cellular stroma, mimicking invasive carcinoma **(c, d)**. Immunohistochemical staining with P63 demonstrates widespread positivity in myoepithelial cells, supporting the diagnosis of fibroadenoma **(e)**.

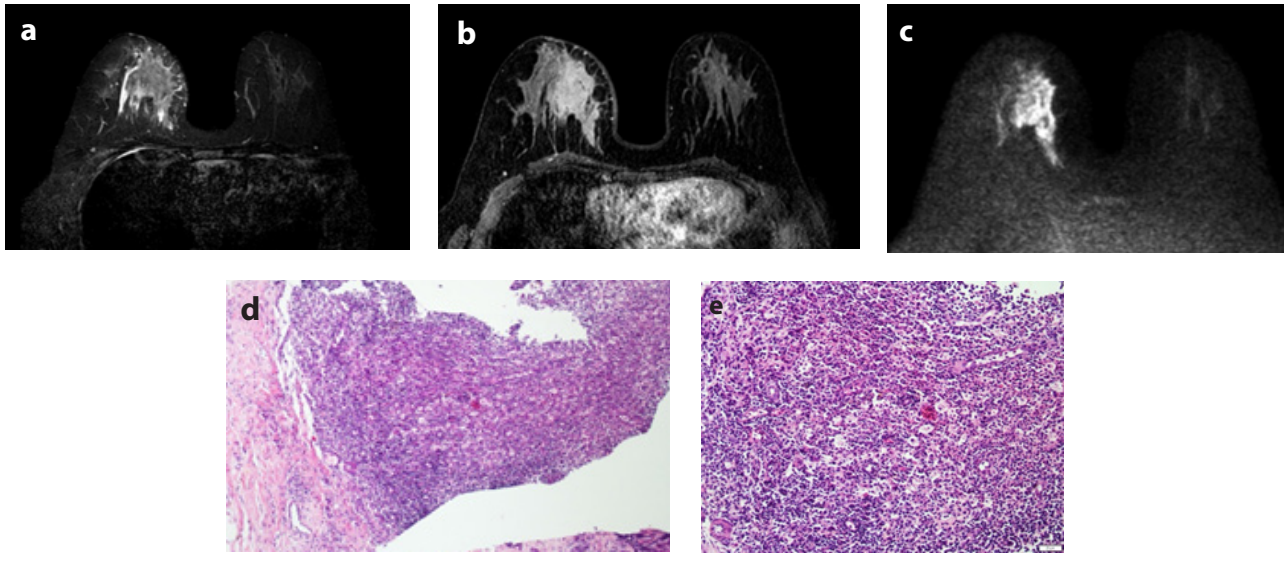

Figure 3. The T2 STIR **(a)**, fat-suppressed post-contrast **(b)**, DWI **(c)**, MRI images of a lesion with spiculated contours and significant diffusion restriction, causing nipple retraction and almost completely filling the lower inner quadrant of the right breast, assessed as BI-RADS 5. Histopathological examination revealed findings of active chronic mastitis accompanied by histiocytes and associated with focal non-necrotizing granuloma formation. (**d, e**; H&Ex10, H&Ex20) *(STIR: Short tau inversion recovery, DWI: Diffusion-weighted imaging, ADC: Apparent diffusion coefficient)*.

the biopsy should be examined through serial sectioning, and if necessary, the recommendation for a repeat biopsy should not be avoided.

In our study, 24 cases of papillary neoplasms were included. A notable finding is that papillary neoplasms constitute 20% of the cases classified as BI-RADS 4B, which represents the highest proportion among the categories. Among the seven cases within this group, three were identified as intraductal papillomas, while the remaining four consisted of papillary proliferations with atypia. Notably, the majority of papillary lesions were classified as BI-RADS 4 (54.2%), followed by BI-RADS 3 as the second most common category (33.3%). All eight cases described as BI-RADS 3 demonstrated histopathological findings consistent with intraductal papilloma upon tru-cut biopsy evaluation. Among the four cases categorized as BI-RADS 4A, three were assessed as intraductal papillomas, while one was interpreted as a papilloma containing DCIS. In the histopathological evaluation of tru-cut biopsies of the two cases defined as BI-RADS 4C, the differential diagnoses included invasive papillary carcinoma and encapsulated carcinoma. The histopathological examination of the excisional materials from these cases revealed a diagnosis of invasive papillary carcinoma accompanied by invasive ductal carcinoma in one case, while the other was diagnosed as invasive ductal carcinoma arising in the periphery of a solid papillary carcinoma. On the other hand, the histopathological evaluation of tru-cut biopsies of the two cases described as BI-RADS 5 primarily indicated intraductal papilloma (excisional materials of these cases could not be obtained) (Fig. 4). Histopathological assessment of papillary neoplasms, which are highly heterogeneous tumors, is open to unexpected findings, and surgical excision is generally recommended.

In our study, fibroadenomas/benign phyllodes tumors were the most common diagnosis among biopsied cases (36%), followed by invasive carcinomas (32%). Patients diagnosed with invasive carcinoma were older, with a mean age of 55.9 years. The rate of malignancy increased as lesion size increased. As expected, the most common quadrant for malignancy was the upper outer quadrant. In this study, we also found that invasive carcinomas tend to occur more frequently in the upper quadrants of the breast, particularly in the upper outer quadrant. Although there are opinions suggesting that the frequent occurrence of malignancies in the upper outer quadrant may be related to cosmetic products, some studies have indicated that the higher malignancy rates in this region could be attributed to the greater amount of breast parenchyma compared to other areas.[16,17] Furthermore, there are publications reporting that genomic instability is higher in the breast parenchyma of the outer quadrants.[18] In our study, the second most common quadrant for malignancy after the upper outer quadrant was the upper inner quadrant. Similarly, in the study conducted by Lee AHS, the upper inner quadrant was also identified as the second most frequent site for malignancy. From this perspective, it can be suggested that masses observed in the upper quadrants have a higher likelihood of being malignant.[16] Cases involving multiple quadrants in this study were most commonly diagnosed with FCC/BBT, as mentioned. Fibrocystic disease can present with various patterns of involvement; however, it is most frequently observed in the upper outer quadrant.[16,19] As mentioned above, this can be explained by the greater amount of breast parenchyma in this region. In our study, FCC/BBT cases were more evenly distributed but showed higher frequencies at specific locations, such as the 9 o'clock position and in diffuse presentations.

The FCC/BBT cases encompass non-neoplastic conditions such as adenosis, granulomatous changes, and inflammatory processes. Nonetheless, in situations where the radiologically identified lesion may not have been fully sampled, there remains a risk that an underlying malignant lesion could have been overlooked. This represents one of the limitations of our study and should be carefully considered when interpreting the results.

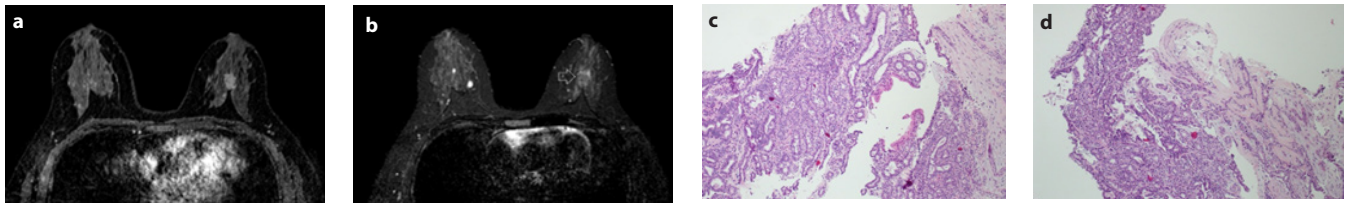

Figure 4. The T2 STIR **(a)** and post-contrast fat-suppressed **(b)** MRI images of a case classified as BI-RADS 5, showing an irregular, partially spiculated lesion with heterogeneous contrast enhancement in the middle quadrant of the right breast. Histopathological examination reveals an intraductal papilloma observed within the duct **(c)**, while hyalinized stroma mimics the appearance of invasive carcinoma **(d)** *(STIR: Short tau inversion recovery).*

CONCLUSION

In routine clinical practice, needle biopsies occupy a pivotal role in the diagnosis of breast tumors, and close collaboration between pathologists and radiologists is of critical importance for accurate diagnosis. There is a remarkably strong concordance between pathology and radiology in evaluations based on the BI-RADS classification system in core biopsies; however, potential diagnostic pitfalls in BI-RADS category 4B and 4C cases must be carefully addressed. In particular, papillary lesions of the breast and inflammatory processes may represent challenging pitfalls in diagnosis.

Currently, innovative studies utilizing advanced methodologies such as deep learning and artificial intelligence are being conducted to further enhance the BI-RADS scoring system.[20,21] The BI-RADS system, which has been effectively employed in the diagnosis of breast lesions both historically and in contemporary practice, is anticipated to undergo further refinements that will enhance its diagnostic utility in the future. As tumors are increasingly characterized at the molecular level—breast tumors perhaps being among the most comprehensively studied—a scoring system that incorporates molecular profiling is emerging as a key expectation for future advancements.[22] Therefore, we propose that collaborative efforts, particularly among radiologists, pathologists, and clinicians, will play a pivotal role in advancing the diagnosis and treatment of these patients.

DECLARATIONS

Ethics Committee Approval: The study was approved by Bağcılar Training and Research Hospital Ethics Committee (No: 2023/08/08/043, Date: 08/08/2023).

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Use of AI for Writing Assistance: Not declared.

Authorship Contributions: Concept – SC, GGP, MEI, AM, ATC; Design – SC, GGP, AM, ATC; Supervision – SC, GGP; Data collection &/or processing – SC, GGP, MEI, OG, EC, MC; Analysis and/or interpretation – SC, GGP, MEI, OG, EC, MY, MC; Literature search – SC, GGP, MEI, OG, EC; Writing – SC, GGP, MEI, OG, MY, AM, ATC; Critical review – SC, GGP.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Burnside ES, Sickles EA, Bassett LW, Rubin DL, Lee CH, Ikeda DM, et al. The ACR BI-RADS experience: learning from history. J Am Coll Radiol 2009;6:851–60.
2. Liberman L, Menell JH. Breast imaging reporting and data system (BI-RADS). Radiol Clin North Am 2002;40:409–30, v.
3. Rao AA, Feneis J, Lalonde C, Ojeda-Fournier H. A pictorial review of changes in the BI-RADS fifth edition. Radiographics 2016;36:623–39.
4. Spak DA, Plaxco JS, Santiago L, Dryden MJ, Dogan BE. BI-RADS® fifth edition: A summary of changes. Diagn Interv Imaging 2017;98:179–90.
5. Tao Z, Shi A, Lu C, Song T, Zhang Z, Zhao J. Breast cancer: epidemiology and etiology. Cell Biochem Biophys 2015;72:333–8.
6. Khan P, Masroor I, Alam MS, Salam A, Ali Y, Khan MS. Sonographic characteristics and pathology correlation of breast imaging reporting and data System (BI-RADS) category 4 lesions. Cureus 2023;15:e51410.
7. Elverici E, Barça AN, Aktaş H, Özsoy A, Zengin B, Çavuşoğlu M, et al. Nonpalpable BI-RADS 4 breast lesions: sonographic findings and pathology correlation. Diagn Interv Radiol 2015;21:189–94.
8. Mohapatra SK, Mishra A, Sahoo TK, Nayak RB, Das PK, Nayak B. The positive predictive values of the breast imaging reporting and data system (BI-RADS) 4 lesions and its mammographic morphological features. Indian J Surg Oncol 2021;12(1):182–9.
9. Yoon JH, Kim MJ, Moon HJ, Kwak JY, Kim EK. Subcategorization of ultrasonographic BI-RADS category 4: positive predictive value and clinical factors affecting it. Ultrasound Med Biol 2011;37:693–9.
10. He P, Cui LG, Chen W, Yang RL. Subcategorization of ultrasonographic BI-RADS category 4: assessment of diagnostic accuracy in diagnosing breast lesions and influence of clinical factors on positive predictive value. Ultrasound Med Biol 2019;45:1253–8.
11. Kim EY, Bae HS, Lee SR, Yun J, Park YL, Park CH. Predictive value of BI-RADS category 4A and 4B lesions detected on breast ultrasonography: single center experience. J Surg Ultrasound 2020;7:55–61.
12. Bent CK, Bassett LW, D'Orsi CJ, Sayre JW. The positive predictive value of BI-RADS microcalcification descriptors and final assessment categories. AJR Am J Roentgenol 2010;194:1378–83.
13. Kim EK, Ko KH, Oh KK, Kwak JY, You JK, Kim MJ, et al. Clinical application of the BI-RADS final assessment to breast sonography in conjunction with mammography. AJR Am J Roentgenol 2008;190:1209–15.
14. Baek SE, Kim MJ, Kim EK, Youk JH, Lee HJ, Son EJ. Effect of clinical information on diagnostic performance in breast sonography. J Ultrasound Med 2009;28:1349–56.
15. Lee HJ, Kim EK, Kim MJ, Youk JH, Lee JY, Kang DR, et al. Observer variability of Breast Imaging Reporting and

Data System (BI-RADS) for breast ultrasound. Eur J Radiol 2008;65:293–8.

16. Lee AH. Why is carcinoma of the breast more frequent in the upper outer quadrant? A case series based on needle core biopsy diagnoses. Breast 2005;14:151–2.
17. Darbre PD. Underarm cosmetics and breast cancer. J Appl Toxicol 2003;23:89–95.
18. Ellsworth DL, Ellsworth RE, Love B, Deyarmin B, Lubert SM, Mittal V, et al. Outer breast quadrants demonstrate increased levels of genomic instability. Ann Surg Oncol 2004;11:861–8.
19. Dang BQ, Miles B, Young P, He J, Nguyen QD. An interesting imaging presentation of a common benign entity: fibrocystic changes in a postmenopausal patient. Cureus 2023;15:e36292.
20. Yang Y, Zhong Y, Li J, Feng J, Gong C, Yu Y, et al. Deep learning combining mammography and ultrasound images to predict the malignancy of BI-RADS US 4A lesions in women with dense breasts: a diagnostic study. Int J Surg 2024;110:2604–13.
21. Haver HL, Yi PH, Jeudy J, Bahl M. Use of ChatGPT to assign BI-RADS assessment categories to breast imaging reports. AJR Am J Roentgenol 2024;223:e2431093.
22. Eghtedari M, Chong A, Rakow-Penner R, Ojeda-Fournier H. Current status and future of BI-RADS in multimodality imaging, from the *AJR* special series on radiology reporting and data systems. AJR Am J Roentgenol 2021;216:860–73.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2025.79847
Eur Arch Med Res 2026;42(2):148–155

Comparison of HALP and Ranson Scores in Predicting Outcomes of Acute Pancreatitis Admitted to the Emergency Department

Ibrahim Altundag,[1] Alperen Sahin,[2] Metin Oner,[1] Kaan Yusufoglu,[1] Burcu Genc Yavuz,[1] Sinem Dogruyol,[1] Mustafa Ahmet Afacan,[1] Sahin Colak[1]

[1]Department of Emergency Medicine, Haydarpasa Numune Training and Research Hospital, Istanbul, Turkiye
[2]Department of Emergency Medicine, Bitlis State Hospital, Bitlis, Turkiye

ABSTRACT

Objective: Acute pancreatitis (AP) often manifests as acute abdominal pain and carries significant risks of mortality and healthcare costs. Its incidence is rising, largely due to alcohol consumption and obesity. Several scoring systems, including the Ranson, Balthazar Severity score, and bedside index for severity in AP, are used to assess AP severity and prognosis, but challenges remain due to the need for prolonged observation and imaging. The present study assesses the ability of the Ranson and HALP scores to predict clinical outcomes and survival in patients with AP.

Materials and Methods: This retrospective analysis included patients with AP, collecting data on demographics, vital signs, laboratory values, comorbidities, and outcomes. The Ranson and HALP scores were calculated, and patients were grouped into survivors and non-survivors based on mortality. Receiver operating characteristic analysis was performed to evaluate the predictive ability of these scores for complications and mortality.

Results: The study included 598 patients, with an overall mortality rate of 12.2%. The non-survivor group had significantly higher mean age and a higher prevalence of malignancy and pancreatitis caused by pancreatic cancer. Severe pancreatitis (Ranson score ≥3) was observed in 21.1% of patients, with these patients exhibiting a higher mortality rate (30.2%). The HALP score demonstrated superior performance over the Ranson score in predicting long-term mortality (area under the curve: 0.800, 95% CI: 0.735–0.864, p=0.000). The first quartile group of the HALP score had significantly higher incidences of complications, severe pancreatitis, and mortality.

Conclusion: The HALP score offers a reliable and rapid alternative to the Ranson score for predicting mortality in AP patients, especially in emergency department settings where timely decision-making is crucial. Its simplicity and predictive value make it a useful tool for clinicians.

Keywords: Acute pancreatitis, ALP score, Emergency department, Mortality, Prognosis, Ranson score

Cite this article as: Altundag I, Sahin A, Oner M, Yusufoglu K, Genc Yavuz B, Dogruyol S, et al. Comparison of HALP and Ranson Scores in Predicting Outcomes of Acute Pancreatitis Admitted to the Emergency Department. Eur Arch Med Res 2026;42(2):148–155.

Address for correspondence: Ibrahim Altundag. Department of Emergency Medicine, Haydarpasa Numune Training and Research Hospital, Istanbul, Turkiye

E-mail: dr.ibrahimaltundag@gmail.com **ORCID ID:** 0000-0002-0880-7218

Submitted: 21.08.2025 **Revised:** 17.09.2025 **Accepted:** 14.10.2025 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

INTRODUCTION

Acute pancreatitis (AP) constitutes a major clinical condition presenting with acute abdominal pain. AP imposes significant physical and psychological distress on patients due to its rapid progression and high mortality rate, while also leading to substantial healthcare expenditures. A recent review estimated that AP accounts for approximately 280,000 hospitalizations annually in the U.S., contributing to healthcare costs of over $2.6 billion.[1] The incidence of AP has increased by approximately 3% over the past 60 years, driven by advancements in imaging techniques and the growing prevalence of risk factors such as alcohol consumption and obesity.[2] Given the potential for severe complications following AP, early risk stratification using reliable scoring systems is essential in clinical practice.[3,4] Mild and self-limiting AP episodes are more common; however, 15–25% of cases progress to moderate or severe forms, resulting in prolonged hospitalization and increased mortality.[5]

Scoring systems are frequently utilized in clinical practice to stratify the severity and predict the prognosis of AP. For this purpose, the Ranson criteria, Balthazar grading system, Acute Physiology and Chronic Health Evaluation II (APACHE II), the computed tomography severity index, and bedside index for severity in AP (BISAP) are widely utilized in clinical practice.[6] Scoring systems assist in guiding patient care and facilitating decisions regarding transfer to intensive care by providing accurate and objective prognosis predictions. Scoring systems have varying sensitivity and specificity values. The Ranson criteria require multiple parameters, some of which are obtained up to 48 h after hospitalization, presenting significant challenges for early evaluation of patients in the emergency department (ED).

The Ranson score has been a key method for assessing the severity of AP for over 30 years; however, it requires 48 h to complete, which significantly limits its applicability in the ED.[6] The HALP score, comprising hemoglobin, albumin, lymphocyte, and platelet counts, was first introduced in 2015 as an index to assess inflammation and nutritional status.[7] Research indicates that the HALP score effectively estimates disease outcomes across different pathologies.[8]

This study evaluated the clinical presentations and laboratory findings of AP patients admitted to the ED. Our aim was to assess the capacity of the Ranson and HALP scores to predict patient outcomes and mortality in AP, alongside exploring other contributing variables.

MATERIALS AND METHODS

This is a retrospective, cross-sectional study. The study analyzed patients diagnosed with AP at the ED of Haydarpaşa Numune Training and Research Hospital Hospital between January 2020 and January 2025. Patients presenting with abdominal pain to the ED were enrolled in this study. The annual number of patient visits to our ED is approximately 235.000, with around 8% of these visits being attributed to abdominal pain.

The diagnosis of AP was established when at least two of the following criteria were met: abdominal pain typical of AP, serum amylase and/or lipase levels more than three times the upper normal limit, or imaging findings characteristic of the disease.[5]

Patients aged ≥18 years with confirmed AP were included in the study. The exclusion criteria for the study were as follows: A history of hospitalization for any reason within the past 30 days, patients who received intravenous drug therapy, the presence of other gastrointestinal disorders, incomplete medical records. Following the diagnosis of AP, patients were managed in accordance with established clinical guidelines.

The data recorded from the hospital information system included age, gender, comorbidities, respiratory rate, systolic and diastolic blood pressure, fever, heart rate, lymphocyte count, neutrophil count, platelet count, hemoglobin, hematocrit, glucose levels, blood urea nitrogen (BUN), aspartate transaminase, alanine transaminase, and C-reactive protein (CRP) levels. In addition, the etiology of pancreatitis, complications, discharge or hospitalization status, length of hospital stay, and mortality were also documented.

Following enrollment, the patients were stratified according to pancreatitis-related mortality: survivors and non-survivors. Non-survivors were further categorized into three groups based on the timing of mortality: Short-term mortality (<1 month), mid-term mortality (1–6 months), and long-term mortality (6–9 months). Using the data from patients' admissions, the CRP/Albumin ratio, Ranson score, and HALP score were calculated. The HALP score was determined for all patients as described below:[7]

HALP = hemoglobin (g/L) × albumin (g/L) × lymphocyte (/L) / platelet (/L)

A Ranson score of 3 or higher was considered indicative of severe pancreatitis, whereas a score below 3 corresponded to mild disease. The demographic characteristics, Ranson score, and HALP score were analyzed across the survivor and non-survivor groups. Subsequently, the predictive value of the Ranson and HALP scores for complications and survival in AP patients was analyzed.

Ethics Statement

This study received approval from the Institutional Ethics Committee (Approval No: HNEAH-GOAEK 2025/53). The study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Because this study was retrospective,

the Clinical Research Ethics Committee of University of Health Sciences Haydarpasa Numune Training and Research Hospital granted a waiver for informed consent.

Statistical Analysis

Statistical analyses were performed using IBM Statistical Package for the Social Sciences 23.0 (IBM Corp., Armonk, NY) and MedCalc 23.1.7. Descriptive statistics are presented as mean±standard deviation (SD) and as percentages. The normality of continuous variables was evaluated using the Kolmogorov–Smirnov test. For continuous variables, normally distributed data are shown as mean±SD, while non-normal data are expressed as median (interquartile range). Associations between categorical variables were examined using the Pearson Chi-square test. However, for categorical variables with expected frequencies below 5, Fisher's exact test was applied to ensure the validity of the results. Comparisons of continuous variables between two groups were conducted using the Student's t-test or Mann–Whitney U-test, as appropriate. The receiver operating characteristic (ROC) analysis was used to evaluate the ability of the scoring systems to predict outcomes. For the HALP score, four quartiles (25th, 50th, and 75th percentiles) were created based on the median values. A $p<0.05$ was considered statistically significant, and results are reported with 95% confidence intervals (CI).

RESULTS

The study group comprised 598 patients who met the inclusion criteria. Of these, 302 (50.5%) were male, with a mean age of 57.59±18.07 years. Twenty-four patients (4%) required intensive care unit admission, and their median length of stay was 5 days, ranging from 1 to 64 days. The overall mortality rate for all patients was 12.2% (73 patients). Among the non-survivor patients, 16 (21.9%) died within the 1st month (short-term), 8 (11.0%) between 1 and 6 months (mid-term), and 49 (67.1%) between 6–9 months (long-term). The non-survivor group had a significantly greater mean age compared with survivors (mean difference: 17.65, 95% CI: 13.84–21.46, $p<0.001$). The comorbidities of the patients were assessed. Malignancy was significantly more common among non-survivors ($p<0.001$). In addition, the etiologies of pancreatitis were examined. The non-survivor group had a significantly higher incidence of pancreatic cancer-related pancreatitis ($p<0.001$). The other etiologies of AP showed no significant differences between the groups. Patient characteristics are summarized in Table 1.

Table 1. Comparison of etiology and demographic characteristics according to outcome groups

	Total (n=598)	Survivor (n=525)	Non-survivor (n=73)	p
Age (years)	57.59 ±18.07	55.43±17.40	73.08±15.07	<0.001
Sex (male)	302 (50.5)	266 (44.5)	36 (6.0)	0.829
Comorbidities, n (%)				
Hypertension	219 (36.6)	188 (31.4)	31 (5.2)	0.269
Diabetes mellitus	135 (22.6)	121 (20.2)	14 (2.3)	0.459
Coronary artery disease	69 (11.5)	60 (10.0)	9 (1.5)	0.822
Hyperlipidemia	52 (8.7)	47 (7.9)	5 (0.8)	0.550
Malignancy (any)	46 (7.7)	31 (5.2)	15 (2.5)	<0.001
COPD	38 (6.4)	29 (4.8)	9 (1.5)	0.026
Chronic kidney disease	36 (6.0)	31 (5.2)	5 (0.8)	0.751
Hypothyroidism	30 (5.0)	24 (4.0)	6 (1.0)	0.181
Congestive heart failure	22 (3.7)	20 (3.3)	2 (0.3)	0.649
Cerebrovascular disease	21 (3.5)	17 (2.8)	4 (0.7)	0.330
Other	54 (9.0)	52 (8.7)	2 (0.3)	0.045
Etiology, n (%)				
Gallstones	350 (58.5)	303 (57.7)	47 (64.4)	0.278
Idiopathic	196 (32.8)	181 (34.5)	15 (20.5)	0.181
Hyperlipidemia	29 (4.8)	26 (5.0)	3 (4.1)	0.753
Alcohol	13 (2.2)	12 (2.3)	1 (1.4)	0.615
Pancreatic malignancy	10 (1.7)	3 (0.6)	7 (9.6)	<0.001

COPD: Chronic obstructive pulmonary disease.

The vital parameters and laboratory findings for the mortality groups are shown in Table 2. Non-survivors had lower diastolic blood pressure (66.99±13.35 vs. 70.95±13.33 mmHg, p=0.021) and higher heart rate (86.67±15.87 vs. 79.73±16.10 bpm, p=0.001) compared to survivors. Oxygen saturation was also significantly lower in non-survivors (93.0 [3.0] vs. 96.0 [3.0]%, p<0.001), while body temperature and systolic blood pressure did not differ significantly. Non-survivors exhibited lower lymphocyte counts, hemoglobin, hematocrit, and albumin levels (all p<0.01), whereas blood glucose, BUN, CRP, and CRP-to-albumin ratio were significantly higher. The HALP score was markedly reduced in non-survivors (1.79 [1.60] vs. 3.57 [2.68], p<0.001), reflecting impaired inflammatory and nutritional status. In addition, non-survivors had higher intensive care unit admission rates (21.9% vs. 1.5%, p<0.001), longer hospital stays (6.0 [6.0] vs. 5.0 [4.0] days, p<0.001), and more frequent complications (32.9% vs. 12.2%, p<0.001) than survivors. AP-related complications were identified in 88 out of 598 patients (14.7%).

According to the Ranson scores, severe pancreatitis (≥3) was detected in 126 patients (21.1%). The mortality rate in these 126 patients with severe pancreatitis was 30.2%, which was significantly higher (p<0.001). The effectiveness of the Ranson and HALP scores in predicting mortality was evaluated using ROC curve analysis (Fig. 1). The Ranson score and HALP score demonstrated similar effectiveness in predicting short-term mortality. The predictive performance of the HALP and Ranson scores for mortality was compared using ROC curve analysis. No significant difference was observed between the two scores for mortality within <1 month (p=0.895) and 1–6 months (p=0.737). However, for long-term mortality at 6–9 months, the HALP score demonstrated a significantly higher area under the curve (AUC) of 0.800 compared to the Ranson score (p=0.000), indicating superior predictive accuracy for long-term outcomes (Table 3).

Table 2. Comparison of vital signs, laboratory results, and outcomes between survivor and non-survivor groups

	Total (n=598)	Survivor (n=525)	Non-survivor (n=73)	p
Vital parameters				
Systolic blood pressure (mmHg)	121.80±18.69	122.34±18.13	117.88±22.03	0.056
Diastolic blood pressure (mmHg)	70.47±13.79	70.95±13.33	66.99±13.35	0.021
Heart rate (beat/minute)	80.58±16.21	79.73±16.10	86.67±15.87	0.001
Oxygen saturation (%)	95.0 (3.0)	96.0 (3.0)	93.0 (3.0)	<0.001
Body temperature (°C)	36.64±0.32	36.64±0.33	36.60±0.26	0.243
Laboratory results				
White blood cell ($\times10^9$/L)	11.35±4.63	11.25±4.47	12.09±5.64	0.145
Lymphocyte ($\times10^9$/L)	1.70±0.82	1.79±0.82	1.10±0.51	<0.001
Neutrophils ($\times10^9$/L)	8.88±4.43	8.77±4.41	9.71±4.49	0.097
Platelet ($\times10^9$/L)	240.2±68.2	241.0±67.4	234.6±73.8	0.480
Hemoglobin (g/dL)	11.99±1.58	12.16±1.48	10.82±1.75	<0.001
Hematocrit (%)	39.51±5.54	39.74±5.39	37.83±6.32	0.006
Blood glucose (mg/dL)	122.0 (55.0)	121.0 (53.0)	134.0 (63.0)	0.001
Blood urea (mg/dL)	14.7 (9.0)	14.0 (8.0)	20.0 (12.0)	<0.001
Aspartate transaminase (U/L)	79 (217)	80 (221)	78 (184)	0.484
Alanine transaminase (U/L)	64 (215)	71 (242)	55 (103)	0.042
Albumin (g/dL)	41.13±4.20	41.79±3.74	36.44±4.37	<0.001
C-reactive protein (mg/L)	1.1 (3.7)	0.9 (3.4)	2.8 (6.8)	0.001
HALP score	3.34 (2.74)	3.57 (2.68)	1.79 (1.60)	<0.001
C-reactive protein-to-albumin ratio	0.25 (0.95)	0.23 (0.83)	0.76 (1.84)	<0.001
Outcomes				
ICU admission (%)	24 (4.0)	8 (1.5)	16 (21.9)	<0.001
LOS (days)	5.0 (4.0)	5.0 (4.0)	6.0 (6.0)	<0.001
Complications (%)	88 (14.7)	64 (12.2)	24 (32.9)	<0.001

Values expressed as number (%), Mean±Standard Deviation, median (IQR: Interquartile Range). ICU: Intensive Care Unit; LOS: Length of the Hospital Stay.

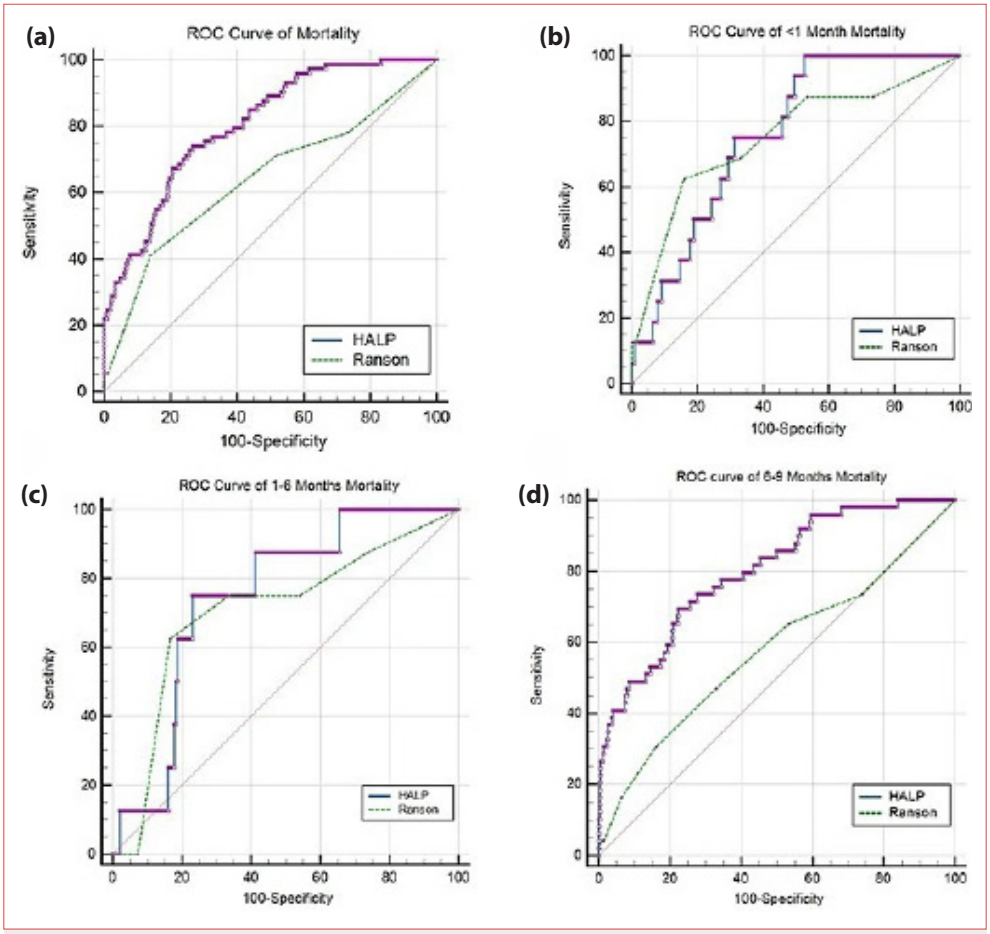

Figure 1. Receiver-operator characteristic curves (ROC curves) of the comparison of HALP and Ranson scores in predicting mortality. **(a)** The ROC curve of overall mortality (p=0.000). **(b)** The ROC curve showing mortality for <1 month (p=0.895). **(c)** The ROC curve showing mortality for 1–6 months (p=0.737). **(d)** The ROC curve showing mortality for 6–9 months (p=0.000).

Table 3. The ROC curve analysis of Ranson and HALP scores in predicting mortality in acute pancreatitis

	Mortality							
	Total (n=73)		<1 months (n=16)		1–6 months (n=8)		6–9 months (n=49)	
	AUC	p	AUC	p	AUC	p	AUC	p
Ranson score	0.641 (CI:0.565–0.717)	<0.001	0.752 (CI:0.615–0.890)	0.001	0.706 (CI:0.512–0.899)	0.046	0.577 (CI:0.484–0.670)	0.074
HALP score	0.804 (CI:0.754–0.854)	<0.001	0.760 (CI:0.673–0.847)	<0.001	0.746 (CI:0.618–0.875)	0.017	0.800 (CI:0.735–0.864)	<0.001

AUC: Area Under the Curve; CI: Confidence Interval.

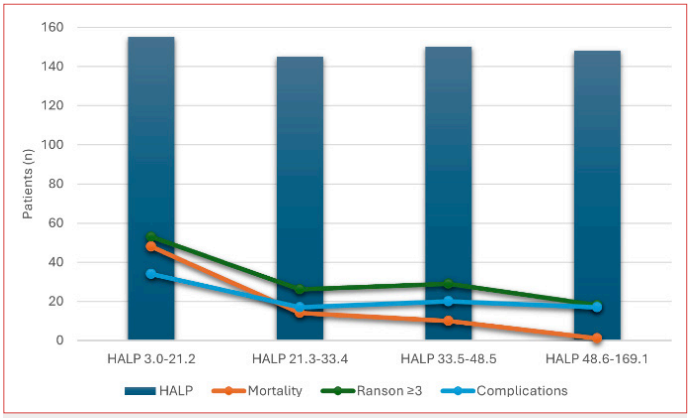

Figure 2. Distribution of patients with severe pancreatitis, complications, and mortality across HALP score quartiles.

The prognostic value of HALP score cutoff points for predicting mortality in AP patients was assessed. For a HALP score lower than 24, sensitivity was 71.2%, specificity was 74.9%, the PPV was 28.3%, and the NPV was 94.9%.

The frequency of complications, severe pancreatitis, and mortality was analyzed based on the 25th, 50th, and 75th percentiles of the HALP score (Fig. 2). The incidence of complications ($p=0.030$), severe AP ($p<0.001$), and mortality rate ($p<0.001$) in the first quartile group of the HALP score (3.0–21.2) were notably higher than in the other quartile groups.

DISCUSSION

We assessed the predictive performance of the HALP and Ranson scores for disease severity and mortality in patients with AP. Our findings indicated that the HALP score is a robust tool for stratifying patients with AP by severity during their initial evaluation in the ED. To date, no studies have compared the predictive performance of HALP and Ranson scores in patients with AP.

Papachristou et al.[6] determined the mortality rate to be 3.8% in their prospective cohort study. Güler et al.[8] reported a mortality rate of 9.1%. In the literature, the mortality rate of complicated pancreatitis has been reported to be between 10% and 30%.[2] In our study, the overall mortality rate was 12%. The mortality rate in patients with complicated AP was 27.3%. The mortality rate was found to be 30.2% in cases of severe AP.

Clinicians must diagnose severe AP early and be attentive during treatment, as severe AP cases are associated with high mortality rates. This has led clinicians to develop scoring systems for the early detection of severe AP, resulting in the creation of numerous different scoring systems. Although the APACHE-II score is among the most commonly used scoring systems for evaluating AP patients in the United States,[6] its application in the ED is limited due to the extensive range of parameters required for evaluation, making it impractical for rapid clinical decision-making. The Ranson score is determined by evaluating 11 parameters, with a score of 3 or more indicating severe disease. However, its reliance on multiple parameters and the necessity of a 48-h waiting period for calculation make it challenging to apply in fast-paced ED settings.

The HALP score has been highlighted as a potential indicator of both inflammatory status and nutritional status in patients.[9] Anemia-related tissue hypoxia can exacerbate inflammatory responses, while thrombosis may further promote inflammation through ischemia-reperfusion mechanisms.[9-11] Lymphocytes play a crucial role in modulating inflammation, and lymphopenia has been linked to adverse outcomes in various inflammatory diseases.[12] Serum albumin levels are an important marker for both nutritional status and the severity of the inflammatory response, with lower levels often associated with greater severity of inflammatory conditions.[13] The HALP score, calculated using only four parameters, offers the advantage of rapid assessment in ED settings compared to the more complex Ranson score. Recently, the HALP score has been utilized to predict prognosis in various patient populations, including those with gastric cancer, prostate cancer, colorectal cancer, pancreatitis, and stroke.[7-9,14,15]

Pancreatic cancer can be an underlying cause of AP and remains a lethal disease with poor prognosis. It leads to significant cancer-related mortality, with an overall 5-year survival rate of about 9%.[16] In our cohort, pancreatic cancer was significantly more frequent among non-survivors. Consequently, AP patients with underlying pancreatic cancer had higher mortality rates, suggesting that poor prognosis may be driven by the malignancy rather than AP alone. Components of the HALP score, such as low hemoglobin and albumin, are also commonly seen in cancer patients. This indicates that the HALP score may, in part, reflect the effect of high-risk comorbidities such as pancreatic cancer when predicting outcomes.

The superior performance of the HALP score in predicting 6–9-month mortality may be explained by its reflection of longer term patient factors, such as nutritional status and chronic inflammation, rather than solely the severity of the AP episode. In contrast, the Ranson score primarily captures acute disease severity, which may explain why it is less predictive of longer-term outcomes.

In a meta-analysis including 5476 AP patients, the AUC for predicting mortality with the Ranson and BISAP scores was found to be 0.91 and 0.92 respectively.[17] A separate study with 634 AP patients reported an AUC of 0.891 (95% CI: 0.833–0.949) for the HALP score in estimating short-term mortality. For a HALP score ≤15 cutoff value, the sensitivity and specificity

were determined to be 82.8% and 86.8%, respectively.[8] In our study, the AUC for predicting mortality using the Ranson and HALP scores were 0.641 and 0.804, respectively. For predicting mortality between 6 and 9 months, the HALP score demonstrated superior performance, with an AUC of 0.800 (95% CI: 0.735–0.864, p<0.001), which was more successful compared to the Ranson score.

Limitations

This research has some inherent limitations. The main limitation of this study is its retrospective, single-center design, which may limit the generalizability of the results. In addition, although the sample size was relatively large, being conducted in a tertiary care ED may have introduced a selection bias toward more severe presentations of AP. Third, certain potential confounding factors – such as variations in treatment protocols, nutritional status prior to admission, and timing of laboratory measurements – could not be fully controlled due to the nature of retrospective data collection. Fourth, the HALP and Ranson scores were calculated based on laboratory values obtained at admission; serial measurements and dynamic changes over time were not assessed, which may have provided additional prognostic information. Fifth, although the HALP score has been associated with prognosis in various malignancy studies, our analysis did not evaluate the subtypes or severity of malignancies in the cohort. This represents a limitation of the study, as differences in malignancy types and severity could potentially influence HALP scores and their predictive performance. External validation in multicenter prospective cohorts is needed to verify the ability of the HALP score to predict outcomes in AP.

CONCLUSION

The results of our study indicate that the HALP score demonstrates a predictive value similar to the Ranson score in assessing mortality risk in patients with AP. The reliability of the Ranson score has been extensively documented in the academic literature. However, its use in ED is limited due to the inclusion of follow-up values obtained 48 h after admission. Given the high workload in ED, the HALP score, which can be calculated in a short time, will serve as a valuable tool for predicting mortality and guiding appropriate treatment planning for patients in the emergency setting.

DECLARATIONS

Ethics Committee Approval: The study was approved by University of Health Sciences Haydarpasa Numune Training and Research Hospital Ethics Committee (No: HNEAH-GOAEK 2025/53, Date: 22/04/2025).

Informed Consent: The Clinical Research Ethics Committee of Haydarpasa Numune Training and Research Hospital granted a waiver for informed consent.

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: Not declared.

Authorship Contributions: Concept – IA, KY, SD; Design – IA, AS, MO, KY, BGY, SD, MAA, SC; Supervision – MAA, SC; Data collection &/or processing – IA, AS, MO, BGY, MAA, SC; Analysis and/or interpretation – IA, AS, KY, SD; Literature search – IA, KY, BGY, SD; Writing – IA, AS, MO, BGY, MAA, SC.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Peery AF, Crockett SD, Murphy CC, Lund JL, Dellon ES, Williams JL, et al. Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: Update 2018. Gastroenterology 2019;156:254–72.e11.
2. Huang Y, Badurdeen DS. Acute pancreatitis review. Turk J Gastroenterol 2023;34:795–801.
3. Xiao AY, Tan ML, Wu LM, Asrani VM, Windsor JA, et al. Global incidence and mortality of pancreatic diseases: a systematic review, meta-analysis, and meta-regression of population-based cohort studies. Lancet Gastroenterol Hepatol 2016;1:45–55.
4. Mederos MA, Reber HA, Girgis MD. Acute pancreatitis: a review. JAMA 2021;325:382–90.
5. Tenner S, Vege SS, Sheth SG, Sauer B, Yang A, et al. American college of gastroenterology guidelines: management of acute pancreatitis. Am J Gastroenterol 2024;119:419–37.
6. Papachristou GI, Muddana V, Yadav D, O'Connell M, Sanders MK, Slivka A, et al. Comparison of BISAP, Ranson's, APACHE-II, and CTSI scores in predicting organ failure, complications, and mortality in acute pancreatitis. Am J Gastroenterol 2010;105:435–42.
7. Chen XL, Xue L, Wang W, Chen HN, Zhang WH, Liu K, et al. Prognostic significance of the combination of preoperative hemoglobin, albumin, lymphocyte and platelet in patients with gastric carcinoma: a retrospective cohort study. Oncotarget 2015;6:41370–82.
8. Güler İ, Ustaalioğlu İ. Predictive power of HALP score in estimating short-term mortality in patients with acute pancreatitis. Ulus Travma Acil Cerrahi Derg 2023;29:1098–102.
9. Tian M, Li Y, Wang X, Tian X, Pei LL, Wang X, Zhang L, et al. The Hemoglobin, Albumin, Lymphocyte, and Platelet (HALP) score is associated with poor outcome of acute ischemic stroke. Front Neurol 2021;11:610318.
10. Barlas RS, Honney K, Loke YK, McCall SJ, Bettencourt-Silva JH, Clark AB, et al. Impact of hemoglobin levels and anemia

on mortality in acute stroke: analysis of UK regional registry data, systematic review, and meta-analysis. J Am Heart Assoc 2016;5:e003019.

11. Kuroda T, Shiohara E, Homma T, Furukawa Y, Chiba S. Effects of leukocyte and platelet depletion on ischemia--reperfusion injury to dog pancreas. Gastroenterology. 1994;107:1125–34.
12. Qi X, Yang F, Huang H, Du Y, Chen Y, Wang M, et al. A reduced lymphocyte ratio as an early marker for predicting acute pancreatitis. Sci Rep 2017;7:44087.
13. Eckart A, Struja T, Kutz A, Baumgartner A, Baumgartner T, Zurfluh S, et al. Relationship of nutritional status, ınflammation, and serum albumin levels during acute ıllness: a prospective study. Am J Med 2020;133:713–22.e7.
14. Guo Y, Shi D, Zhang J, Mao S, Wang L, Zhang W, et al. The Hemoglobin, Albumin, Lymphocyte, and Platelet (HALP) score is a novel significant prognostic factor for patients with metastatic prostate cancer undergoing cytoreductive radical prostatectomy. J Cancer 2019;10:81–91.
15. Jiang H, Li H, Li A, Tang E, Xu D, Chen Y, et al. Preoperative combined hemoglobin, albumin, lymphocyte and platelet levels predict survival in patients with locally advanced colorectal cancer. Oncotarget 2016;7:72076–83.
16. Rawla P, Sunkara T, Gaduputi V. Epidemiology of pancreatic cancer: global trends, etiology and risk factors. World J Oncol 2019;10:10–27.
17. Zhu J, Wu L, Wang Y, Fang M, Liu Q, Zhang X. Predictive value of the Ranson and BISAP scoring systems for the severity and prognosis of acute pancreatitis: A systematic review and meta-analysis. PLoS One 2024;19:e0302046.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2025.35682
Eur Arch Med Res 2026;42(2):156–161

Neoadjuvant Pertuzumab in Breast Cancer and Associated Prognostic Factors

Orcun Can,[1] Seval Ay,[2] Muhammed Mustafa Atci[3]

[1]Department of Medical Oncology, Acibadem Maslak Hospital, Istanbul, Turkiye
[2]Department of Medical Oncology, Health Science University, Kartal Dr. Lütfi Kirdar City Hospital, Istanbul, Turkiye
[3]Department of Medical Oncology, Health Science University, Prof. Dr. Cemil Tascioglu City Hospital, Istanbul, Turkiye

ABSTRACT

Objective: This study aimed to evaluate the pathological complete response (pCR) rates and predictors of pCR in human epidermal growth factor receptor-2 (HER2)-positive, high-risk locally advanced breast cancer (BC) patients treated with neoadjuvant pertuzumab and trastuzumab combined with chemotherapy in a real-world setting.

Materials and Methods: A retrospective cohort analysis was conducted on 116 patients diagnosed between April 2015 and April 2021. Pathological response was assessed using the Miller-Payne grading system, with pCR defined as the absence of invasive tumor cells in the breast and axillary lymph nodes. Univariate and multivariate analyses were performed to identify predictors of pCR.

Results: The pCR rate was 71.6%, significantly higher than rates reported in pivotal trials. Patients achieving pCR had superior 3-year event-free survival (85% vs. 58%, hazard ratio [HR]: 0.42; *p*=0.002). Younger age (HR: 0.93, *p*=0.03), higher tumor grade (HR: 0.31, *p*=0.016), androgen receptor positivity (HR: 0.41, *p*<0.001), and right-sided tumors (HR: 0.23, *p*=0.026) were independently associated with pCR.

Conclusion: Dual HER2 blockade with pertuzumab and trastuzumab in neoadjuvant therapy yields high pCR rates and improved survival outcomes in HER2-positive BC. Tumor biology and location may influence treatment response, supporting personalized therapeutic strategies.

Keywords: Event-free survival, Human epidermal growth factor receptor-2 positive breast cancer, Neoadjuvant therapy, Pathological complete response, Pertuzumab, Predictive factors

Cite this article as: Can O, Ay S, Atci MM. Neoadjuvant Pertuzumab in Breast Cancer and Associated Prognostic Factors. Eur Arch Med Res 2026;42(2):156–161.

INTRODUCTION

Breast cancer (BC) represents the most frequently diagnosed malignancy and leading cause of cancer deaths in female populations worldwide.[1] A significant proportion of patients (approximately 95%) present with localized or locally advanced disease at initial diagnosis, highlighting the importance of early detection and effective treatment strategies.[2] Among BC subtypes, human epidermal growth factor receptor-2

Address for correspondence: Orcun Can. Department of Medical Oncology, Acibadem Maslak Hospital, Istanbul, Türkiye
E-mail: drorcuncan@gmail.com **ORCID ID:** 0000-0002-9400-105X

Submitted: 07.07.2025 **Revised:** 13.07.2025 **Accepted:** 20.10.2025 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

(HER2)-positive tumors, characterized by HER2 gene amplification or protein overexpression, account for 15–20% of cases and are associated with aggressive tumor behavior, higher metastatic potential, and poorer survival outcomes compared to HER2-negative disease.[3]

The introduction of pre-operative systemic therapy (neoadjuvant therapy, neo-adjuvant chemotherapy [NACT]) has revolutionized the management of BC. Initially, NACT was used primarily to downstage tumors in patients with inoperable disease, rendering them eligible for surgery. However, clinical trials have since demonstrated that NACT not only increases the likelihood of breast-conserving surgery but also provides comparable survival outcomes to adjuvant therapy.[4,5] Trials found associations between postneoadjuvant pathological response and prognosis.[6,7] In contemporary oncology practice, pathological complete response (pCR) – the histological absence of invasive tumor cells in breast parenchyma and axillary nodal tissue after neoadjuvant systemic therapy – has emerged as a validated biomarker predicting superior long-term survival, with particular clinical relevance for HER2-positive and basal-like BC phenotypes.[8,9] Persistent activation of signaling pathways due to amplification of HER2 in BC patients with HER2 positivity leads to a biologically aggressive malignancy with higher susceptibility to cytotoxic chemotherapy. Compared with most other BC subtypes, a higher percentage of HER2-positive patients obtain pCR with NACT even in the absence of HER2-targeted therapy.[6,10] Adding trastuzumab to the neoadjuvant treatment in HER2-positive BC has shown to be beneficial in terms of pCR rate, event-free survival (EFS), and overall survival (OS) in several randomized studies and meta-analyses.[9,11,12]

Pertuzumab, which is a monoclonal antibody, binds to HER2 than trastuzumab and prevents the formation of HER2/HER3 heterodimers, which is the resistance mechanism against trastuzumab. The addition of pertuzumab to NACT and trastuzumab for patients having HER2-positive locally advanced, early-stage (lymph node-positive or >2 cm) BC was approved in 2013.[13] The addition of pertuzumab to the treatment increases the frequency of diarrhea and hematological toxicities. Pertuzumab should be routinely added to the patients treated with NACT and trastuzumab, since pertuzumab increases pathological responses. Clinical trials such as NeoSphere and TRYPHAENA established that combining trastuzumab and pertuzumab as dual HER2-targeted therapy markedly improves pCR rates, which subsequently supported regulatory approval for treating high-risk, early-stage, and locally advanced HER2-positive BC.[14,15] However, this treatment regimen is linked to higher toxicity, particularly gastrointestinal effects such as diarrhea and hematologic side effects, which may restrict its suitability for patients with lower-risk disease.[9,11]

Despite these advances, real-world data on the efficacy of pertuzumab-containing NACT regimens and factors influencing pCR remain limited. This research investigates the rates of pathological response and identifies predictive factors for achieving pCR in HER2-positive, locally advanced BC patients with high-risk features receiving neoadjuvant therapy combining trastuzumab and pertuzumab under real-world clinical conditions.

MATERIALS AND METHODS

Study Population

We conducted a retrospective analysis of medical records from patients diagnosed with HER2-positive BC between April 2015 and April 2021. The study population included individuals with locally advanced, inflammatory, or early-stage HER2+ disease identified through hospital databases. HER2 positivity was determined by immunohistochemistry (IHC 3+ score) or fluorescence in situ hybridization (for cases with IHC 2+ results). Additional inclusion criteria required patients to have an Eastern Cooperative Oncology Group performance status (ECOG PS) of 0–1 and baseline left ventricular ejection fraction ≥55%. Exclusion criteria comprised age under 18 years, non-HER2+ tumor histology, metastatic disease (Stage IV), prior chemotherapy for metastatic cancer, presence of secondary malignancies, renal impairment, hepatic or cardiac dysfunction, and uncontrolled hypertension.

This is a retrospective cohort study. Data of patients were obtained from two high-volume hospitals. Age, hormone status, ECOG PS, Ki-67, tumor grade, tumor location, clinical tumor size, and regional lymph node status were obtained from the medical records of patients. The primary endpoint is pathological evaluation of tumors according to the Miller-Payne grading (MPG) system.[16] MPG is a pathological evaluation, comparing the tumor tissue at the diagnosis with the tumor tissue of the patient who was operated after neoadjuvant chemotherapy. MPG consists of 5 grades whereas grade 1 is pathological no response (pNR) to chemotherapy; grade 2, 3, and 4 are pathological partial response (pPR); and grade 5 is pCR, which is the absence of invasive tumor in the breast and axilla. In the present study, pathological responses are divided into two groups as pCR and non-pCR (pNR+pPR).

Treatment Protocols

Adriamycin 60 mg/m^2 plus cyclophosphamide 600 mg/m^2 intravenously (i.v.) was repeated every 14 days; treatment was completed in four cycles. Sequential paclitaxel (T) 80 mg/m^2 weekly i.v. was repeated every 21 days, trastuzumab was given 8 mg/kg (cycle 1), followed by 6 mg/kg, and plus pertuzumab loading dose was 840 mg (cycle 1), followed by 420 mg and was repeated every 21 days. Treatment was completed in four cycles (4AC plus thermoplastic polyolefin).

Ethics Approval

This investigation was performed in compliance with the Declaration of Helsinki, with ethical clearance granted by the Institutional Review Board of the Göztepe Prof. Dr. Süleyman Yalçın City Hospital Ethics Committee (reference ID:E-23898784-604.01.01).

Statistical Analysis

Continuous variables were analyzed using the Mann–Whitney U test, while categorical variables were assessed through Chi-square tests or Fisher's exact test, as appropriate. The primary study outcome was tumor pathological response evaluated according to the Miller-Payne grading system following pertuzumab-enhanced neoadjuvant chemotherapy. Disease-free survival was defined as the interval from initial diagnosis to either confirmed disease recurrence, mortality, or final clinical assessment. OS was measured from diagnosis until death or the last documented follow-up. Predictive factors for pCR were examined using Cox proportional hazards regression modeling. A 95% confidence interval was employed for all analyses, with statistical significance set at $p<0.05$. All computations were executed using the Statistical Package for the Social Sciences statistical software (version 27.0; IBM Corporation, Armonk, NY).

RESULTS

Patients' Characteristics

From April 2015 to April 2021, 116 patients diagnosed with HER2 + BC receiving neoadjuvant chemotherapy treatment were retrospectively analyzed. All patients had received the same neoadjuvant chemotherapy protocol. Median age was 51 years (26–77). 35.3% of patients were premenopausal, 46.6% were perimenopausal, and 18.1% were postmenopausal. Distribution of molecular subtypes was 62 patients (53.4%) were estrogen receptor (ER) positive or partial response (PR) positive, or both and 54 patients (46.6%) were ER negative and PR negative. Disease type was 15 patients (12.9%) were operable, 96 patients (82.8%) were locally advanced, and 5 patients (4.3%) were inflammatory. Clinical tumor sizes were 23.3% cT1, 48.3% cT2, 22.4% cT3, and 6% cT4. Lymph node status was 56.9% cN1, 38.7% cN2, and 4.4% cN3. 12 patients were assessed as G1, 48 patients were G2, and 56 patients were G3. 59 (50.9%) patients had a right tumor, whereas 57 (49.1%) patients had left side tumor. Patients were divided into two groups according to the MPG system: pCR and non-pCR. 83 patients (71.6%) were in the pCR group, and 33 patients (28.4%) were in non-pCR group. The patients' clinical characteristics were summarized in Table 1.

Table 1. Patient's characteristics

	n (%)
Age (years)	51 (26–77)
Menstrual status	
Premenopausal	41 (35.3)
Perimenopausal	54 (46.6)
Postmenopausal	21 (18.1)
Molecular type	
ER positive or PR positive, or both	62 (53.4)
ER negative and PR negative	54 (46.6)
Disease type	
Operable	15 (12.9)
Locally advanced	96 (82.8)
Inflammatory	5 (4.3)
Clinical tumor size	
cT1	27 (23.3)
cT2	56 (48.3)
cT3	26 (22.4)
cT4	7 (6)
Lymph node status	
cN1	66 (56.9)
cN2	45 (38.7)
cN3	5 (4.4)
Tumor grade	
G1	12 (10.3)
G2	48 (41.4)
G3	56 (48.3)
Tumor location	
Right	59 (50.9)
Left	57 (49.1)
Miller-Payne Grading	
pCR	83 (71.6)
non-pCR	33 (28.4)

pCR: Pathological complete response; non-CR: pPR+pNR: Non-complete response: Pathological partial response + pathological no response.

Patients achieving pCR exhibited superior outcomes, with a 3-year EFS rate of 85% (95% confidence interval [CI]: 76–92%) compared to 58% (95% CI: 45–70%) in the non-pCR group (log-rank *p*<0.001). The hazard ratio (HR) for events (recurrence, metastasis, or death) was 0.42 (95% CI: 0.24–0.73; *p*=0.002), indicating a 58% reduction in risk for the pCR cohort. The survival curves diverged early, with non-pCR patients showing a steeper decline in EFS within the first 24 months.

Predictors of Pathological Response

Univariate analysis identified tumor grade (*p*=0.018), E-cadherin expression (*p*<0.001), androgen receptor (AR) status (*p*<0.001), and tumor location (*p*<0.001) as significant factors influencing pCR (Table 2). Multivariate Cox regression confirmed that younger age (HR: 0.93, *p*=0.03), high-

Table 2. Univariate analysis of pathological response to neoadjuvant chemotherapy

	PCR (%)	non-PCR (%)	p
Age (years)	50 (26–77)	58 (30–77)	0.099
Menopause status			
Premenopausal	32 (78)	9 (22)	0.156
Perimenopausal	17 (80)	4 (20)	
Postmenopausal	34 (62.9)	20 (37.1)	
Grade (G)			
G1	12 (100)	0	0.018
G2	29 (60.4)	19 (39.6)	
G3	42 (75)	14 (25)	
Ki67	40 (13–80)	30 (10–75)	0.834
E-cadherin			
Positive	32 (55.1)	26 (44.9)	0.000
Negative	31(81.5)	7 (18.5)	
Androgen receptor			
Positive	70 (88.6)	9 (11.4)	0.000
Negative	13 (35.1)	24 (64.9)	
Lenfo-vascular invasion			
Positive	32 (78)	9 (22)	0.111
Negative	49 (65.3)	26 (34.7)	
Neural invasion			
Positive	27 (69.2)	12 (30.8)	0.584
Negative	56 (72.7)	21 (27.3)	
Molecular type			
ER positive or PR positive, or both	13 (24.1)	20 (37.3)	0.222
ER negative and PR negative	41 (75.9)	42 (62.7)	
Operable			
Locally advanced	72 (72.9)	25 (27.1)	0.327
Inflammatory	11 (57.8)	8 (42.2)	
Clinical tumor size			
cT1	16 (59.2)	11 (40.8)	0.087
cT2	44 (78.5)	12 (21.5)	
cT3	20 (76.9)	6 (23.1)	
cT4	3 (42.9)	4 (57.1)	
Lymph node status			
cN1	49 (74.2)	17 (25.8)	0.560
cN2	29 (70.7)	12 (29.3)	
cN3	4 (50)	4 (50)	
Tumor location			
Right	20 (48.8)	21 (51.2)	0.000
Left	63 (84)	12 (16)	

PCR: Pathological complete response; ER: Estrogen receptor; PR: Partial response.

Table 3. Multivariate analysis of pathological response to neoadjuvant chemotherapy

	HR	95% CI Minimum-Maximum	p
Age at diagnosis	0.931	0.008–0.993	0.030
Grade	0.031	0.002–0.529	0.016
E-cadherin	1.580	0.412–6.062	0.996
Androgen receptor	0.041	0.007–0.232	<0.001
Clinical tumor size	0.474	0.038–5.850	0.560
Tumor location	0.234	0.065–0.843	0.026

HR: Hazard ratio; CI: Confidence interval.

er tumor grade (HR: 0.31, *p*=0.016), AR positivity (HR: 0.41, *p*<0.001), and right-sided tumors (HR: 0.23, *p*=0.026) were independently associated with pCR (Table 3).

DISCUSSION

This study evaluated the efficacy of neoadjuvant trastuzumab and pertuzumab combined with chemotherapy in 116 HER2-positive BC patients, demonstrating a pCR rate of 71.6%, which is notably higher than rates reported in pivotal trials such as NeoSphere (45.8%) and TRYPHAENA (51.9%).[14,15] The 3-year EFS advantage in pCR patients (85% vs. 58%, HR: 0.42; *p*=0.002) reinforces the prognostic value of pCR, aligning with the CTNeoBC pooled analysis.[6] These findings underscore the clinical benefit of dual HER2 blockade in high-risk HER2-positive BC, particularly in real-world settings where patient selection and treatment adherence may differ from controlled trials.

The higher pCR rate observed in our cohort may reflect the exclusive use of anthracycline-taxane-based chemotherapy combined with dual HER2 blockade, a regimen supported by the BERENICE trial.[17] Notably, our pCR rates approach those reported in the JBCRG-20 study (71% with T-DM1/pertuzumab),[18] suggesting that optimized chemotherapy backbones enhance trastuzumab/pertuzumab efficacy. However, the lack of survival data due to short follow-up limits our ability to correlate pCR with long-term outcomes, a critical gap addressed in trials like KATHERINE, where adjuvant T-DM1 improved survival in residual disease post-neoadjuvant therapy.[19]

Predictors of pCR identified in our study – younger age, higher tumor grade, AR positivity, and right-sided tumors – offer insights into patient selection. The association between AR positivity and pCR (HR: 0.41; *p*<0.001) contrasts with some studies,[20,21] possibly due to differences in tumor biology or treatment protocols. The right-sided tumor advantage (HR:

0.23; *p*=0.026) is a novel finding warranting validation, as prior research focused on prognosis by quadrant rather than response.

Limitations include the retrospective design, small sample size, and heterogeneity in tumor characteristics. The absence of central HER2 testing may have introduced variability, though all cases met standard IHC/FISH criteria. In addition, toxicity data were not systematically collected, precluding safety comparisons with trials like TRYPHAENA, where diarrhea and hematologic toxicities were common.[15]

Future directions should prioritize prospective studies with longer follow-up to assess survival and validate predictive biomarkers. The role of novel agents like trastuzumab deruxtecan in the neoadjuvant setting (e.g., DESTINY-Breast11 trial NCT05113251) could further redefine treatment paradigms.

CONCLUSION

Our real-world findings validate the clinical effectiveness of pertuzumab-based treatment protocols while underscoring the importance of tailored therapeutic approaches guided by individual tumor characteristics. Future advancements should focus on establishing uniform pCR evaluation criteria and incorporating novel treatment modalities to further enhance clinical outcomes in HER2-positive BC management.

DECLARATIONS

Ethics Committee Approval: The study was approved by Göztepe Prof. Dr. Süleyman Yalçın City Hospital Ethics Committee (No: E-23898784-604.01.01, Date: 02/09/2021).

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: Not used.

Authorship Contributions: Concept – OC, SA Design – OC, SA, MMA; Supervision – SA, MMA; Fundings – OC, SA; Materials – OC, SA, MMA; Data collection &/or processing – OC, SA, MMA; Analysis and/or interpretation – SA, MMA; Literature search – OC, SA; Writing – OC; Critical review – SA, MMA.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Siegel RL, Miller KD, Fuchs HE, Jemal A. Cancer statistics, 2021. CA Cancer J Clin 2021;71:7–33.
2. Caswell-Jin JL, Plevritis SK, Tian L, Cadham CJ, Xu C, Stout NK, et al. Change in survival in metastatic breast cancer with treatment advances: meta-analysis and systematic review. JNCI Cancer Spectr 2018;2:pky062.
3. Noone AM, Cronin KA, Altekruse SF, Howlader N, Lewis DR, Petkov VI, et al. Cancer incidence and survival trends by subtype using data from the surveillance epidemiology and end results program, 1992-2013. Cancer Epidemiol Biomarkers Prev 2017;26:632–41.
4. Fisher B, Bryant J, Wolmark N, Mamounas E, Brown A, Fisher ER, et al. Effect of preoperative chemotherapy on the outcome of women with operable breast cancer. J Clin Oncol 1998;16:2672–85.
5. Mauri D, Pavlidis N, Ioannidis JP. Neoadjuvant versus adjuvant systemic treatment in breast cancer: a meta-analysis. J Natl Cancer Inst 2005;97:188–94.
6. Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, et al. Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. Lancet 2014;384:164–72. Erratum in: Lancet 2019;393:986.
7. von Minckwitz G, Untch M, Blohmer JU, Costa SD, Eidtmann H, Fasching PA, et al. Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. J Clin Oncol 2012;30:1796–804.
8. Takada M, Ishiguro H, Nagai S, Ohtani S, Kawabata H, Yanagita Y, et al. Survival of HER2-positive primary breast cancer patients treated by neoadjuvant chemotherapy plus trastuzumab: a multicenter retrospective observational study (JBCRG-C03 study). Breast Cancer Res Treat 2014;145:143–53.
9. Yuan Y, Long X, Wei M, Chen L, Zhang J, Liu X. Long and short-term efficacy and safety comparison of nab-paclitaxel versus paclitaxel combined with trastuzumab and pertuzumab for neoadjuvant treatment of HER2-positive breast cancer: A systematic review and meta-analysis. Cancer Treat Rev 2025;138:102975.
10. Esserman LJ, Berry DA, DeMichele A, Carey L, Davis SE, Buxton M, et al. Pathologic complete response predicts recurrence-free survival more effectively by cancer subset: results from the I-SPY 1 TRIAL--CALGB 150007/150012, ACRIN 6657. J Clin Oncol 2012;30:3242–9.
11. Gianni L, Pienkowski T, Im YH, Roman L, Tseng LM, Liu MC, et al. Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. Lancet Oncol 2012;13:25–32.
12. Akturk Esen S, Seven I, Akdere Ates G, Colak R, Kinikoglu O, Akay Hacan B, et al. Real-life data on adjuvant trastuzumab emtansine treatment in early-stage HER2-positive breast cancer: a Turkish Oncology Group study. Anticancer Drugs 2025;36:575–82.

13. Amiri-Kordestani L, Wedam S, Zhang L, Tang S, Tilley A, Ibrahim A, et al. First FDA approval of neoadjuvant therapy for breast cancer: pertuzumab for the treatment of patients with HER2-positive breast cancer. Clin Cancer Res 2014;20:5359–64.
14. Gianni L, Pienkowski T, Im YH, Tseng LM, Liu MC, Lluch A, et al. 5-year analysis of neoadjuvant pertuzumab and trastuzumab in patients with locally advanced, inflammatory, or early-stage HER2-positive breast cancer (NeoSphere): a multicentre, open-label, phase 2 randomised trial. Lancet Oncol 2016;17:791–800.
15. Schneeweiss A, Chia S, Hickish T, Harvey V, Eniu A, Waldron-Lynch M, et al. Long-term efficacy analysis of the randomised, phase II TRYPHAENA cardiac safety study: Evaluating pertuzumab and trastuzumab plus standard neoadjuvant anthracycline-containing and anthracycline-free chemotherapy regimens in patients with HER2-positive early breast cancer. Eur J Cancer 2018;89:27–35.
16. Ogston KN, Miller ID, Payne S, Hutcheon AW, Sarkar TK, Smith I, et al. A new histological grading system to assess response of breast cancers to primary chemotherapy: prognostic significance and survival. Breast 2003;12:320–7.
17. Swain SM, Ewer MS, Viale G, Delaloge S, Ferrero JM, Verrill M, et al. Pertuzumab, trastuzumab, and standard anthracycline- and taxane-based chemotherapy for the neoadjuvant treatment of patients with HER2-positive localized breast cancer (BERENICE): a phase II, open-label, multicenter, multinational cardiac safety study. Ann Oncol 2018;29:646–53.
18. Komorowski AS, Pezo RC. Untapped "-omics": the microbial metagenome, estrobolome, and their influence on the development of breast cancer and response to treatment. Breast Cancer Res Treat 2020;179:287–300.
19. Modi S, Saura C, Yamashita T, Park YH, Kim SB, Tamura K, et al. Trastuzumab deruxtecan in previously treated HER2-positive breast cancer. N Engl J Med 2020;382:610–21.
20. Lv Y, Li Y, Mu W, Fu H. Factors affecting pathological complete response after neoadjuvant chemotherapy in operable primary breast cancer. J Coll Physicians Surg Pak 2020;30:389–93.
21. Conforti F, Nekljudova V, Sala I, Ascari R, Solbach C, Untch M, et al. Surrogate end points for overall survival in neoadjuvant randomized clinical trials for early breast cancer. J Clin Oncol 2025;43:1441–52.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2025.89266
Eur Arch Med Res 2026;42(2):162–168

Comparison of the Quality of English and Turkish Youtube Videos Related to Pes Equinovarus

Tahsin Olgun Bayraktar, Osman Sariyar, Serdar Aki, Nazim Erkurt, Mehmet Selcuk Saygili, Hakan Gurbuz

Department of Orthopedic and Traumatology, Prof. Dr. Cemil Tascioglu City Hospital, Istanbul, Turkiye

ABSTRACT

Objective: Pes equinovarus is a common congenital foot deformity for which patients and families frequently seek information on social media platforms, such as YouTube. This study aimed to evaluate and compare the quality, reliability, and educational value of English- and Turkish-language YouTube videos related to pes equinovarus.

Materials and Methods: On May 16, 2025, systematic searches were conducted on YouTube using standardized English ("clubfoot," "congenital talipes equinovarus," "talipes equinovarus") and Turkish ("çarpık ayak," "doğumsal çarpık ayak," "talipes equinovarus") terms. The first 150 videos for each language were screened according to pre-defined inclusion and exclusion criteria. After eliminating duplicates, non-relevant, or poor-quality content, 65 English and 51 Turkish videos were included. Video characteristics (duration, views, likes, source, content type) were recorded. Reliability and educational quality were assessed using the Journal of the American Medical Association (JAMA) benchmarks, the global quality score (GQS), the DISCERN instrument, and a novel 20-item clubfoot-specific score (CSS). Assessments were independently performed by two pediatric orthopedic surgeons, with intra- and interobserver reliability tested.

Results: Turkish videos were more often uploaded by physicians (80% vs. 55%, p=0.04), whereas English videos were more frequently produced by research groups or patients/families. Turkish videos had fewer views and likes but demonstrated higher educational quality. Intra- and interobserver reliability was excellent for DISCERN, good for CSS, and moderate for JAMA and GQS. Correlation analysis showed strong positive associations between CSS and DISCERN (r=0.970, p<0.001), CSS and GQS (r=0.695, p<0.001), and a moderate correlation between CSS and JAMA (r=0.410, p=0.032).

Conclusion: The overall quality of YouTube videos on pes equinovarus remains low. Turkish-language content, despite lower popularity, demonstrated higher educational value, likely due to greater physician authorship. The CSS showed strong reliability and may serve as a useful tool for evaluating disease-specific online educational resources.

Keywords: Clubfoot, Health information, Pes equinovarus, Video quality, YouTube

Cite this article as: Bayraktar TO, Sariyar O, Aki S, Erkurt N, Saygili MS, Gurbuz H. Comparison of the Quality of English and Turkish Youtube Videos Related to Pes Equinovarus. Eur Arch Med Res 2026;42(2):162–168.

Address for correspondence: Tahsin Olgun Bayraktar. Department of Orthopedic and Traumatology, Prof. Dr. Cemil Tascioglu City Hospital, Istanbul, Turkiye

E-mail: olgun.bayraktar@yahoo.com **ORCID ID:** 0000-0001-6544-1710

Submitted: 20.08.2025 **Revised:** 27.10.2025 **Accepted:** 30.10.2025 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

INTRODUCTION

Pes equinovarus is one of the most common congenital anomalies of the lower extremity and affects approximately 1–2 newborns/1,000 births.[1,2] This severe foot deformity often causes anxiety among parents and frequently leads them to seek information regarding the condition and treatment options via the Internet. Previous studies have demonstrated that online health information influences patient decision-making.[3-5] YouTube has become a widely used platform due to its visual and auditory content.[6] However, academic content uploaded to YouTube is not peer-reviewed, which may expose users to inaccurate or misleading information.[7,8]

The use of the Internet and social media varies across geographical, linguistic, and cultural contexts, potentially influencing both information-seeking users and content creators.[9] This study sought to evaluate and compare the quality, reliability, and accuracy of pes equinovarus–related YouTube videos presented in Turkish and English. Similar to previous studies in the literature, it was hypothesized that the overall quality of videos would be low, and that academic authorship of content might improve quality and reliability.

MATERIALS AND METHODS

This study was conducted in accordance with the principles of the Declaration of Helsinki. A systematic search of the YouTube™ database was performed on May 16, 2025, using Google Chrome (version 92.0.4515.159, 64-bit) with cleared cache and deleted cookies. The search terms "clubfoot," "congenital talipes equinovarus," and "talipes equinovarus" were used. In the literature, studies analyzing YouTube videos commonly include the first 50 videos for each search term, as these videos represent the most accessed content and the most relevant results according to YouTube's algorithm.[10] Limiting the sample to 50 videos also makes the analysis manageable and comparable. In our study, for each search term, the first 50 videos ranked by YouTube's relevance algorithm were screened, yielding a total of 150 videos.[7,11]

Inclusion criteria were videos in English, primarily related to clubfoot, and with acceptable audiovisual quality. Exclusion criteria were duplicate videos, videos containing only audio or only visual content, YouTube Shorts, non-English videos, irrelevant videos, news, drama, satire, and advertisements unrelated to clubfoot. Multi-part videos were considered as a single entry, and no restriction was applied regarding video length. For the purposes of this study, a YouTube™ account was created, and all video links were recorded after eliminating duplicates. A total of 34 duplicate videos, 1 non-English video, 38 YouTube Shorts, 4 irrelevant videos, and 8 unrelated advertisement videos were excluded. The remaining 65 videos were included in the study.

On the same date, a second standard search was conducted using the Turkish keywords "çarpık ayak," "doğumsal çarpık ayak," and "talipes equinovarus." The same methods were applied, and the first 50 videos for each term (total of 150) were screened. Inclusion criteria were videos in Turkish, primarily related to pes equinovarus, and with acceptable audiovisual quality. Exclusion criteria were identical to the English search. A total of 47 duplicate videos, 4 non-Turkish videos, 16 YouTube Shorts, 23 irrelevant videos, and 9 unrelated advertisement videos were excluded. The remaining 51 videos were included in the study.

For each included video, the following characteristics were recorded: (1) Title, (2) duration, (3) number of views, (4) source/uploader, (5) type of content, (6) number of days since upload, and (7) number of likes. Uploaders were categorized as follows: (1) Physician (affiliated with research groups, universities, or colleges), (2) non-physician healthcare professional, (3) medical source (health websites or animations), (4) patient or family, and (5) advertisement related to clubfoot. The content type was categorized as: (1) Disease-specific information, (2) patient/family experience, (3) treatment techniques and approaches, and (4) advertisement.

Video reliability was assessed using the Journal of the American Medical Association (JAMA) benchmarks.[11,12] This system assigns one point for each of four criteria, with a maximum score of 4 indicating high reliability and a minimum score of 0 indicating low reliability. Although not validated, these criteria have been widely applied in the literature for evaluating online health information.[12,13]

Educational value was assessed using three different scoring systems. The global quality score (GQS) provides a five-point scale to assess overall educational quality.[14-16] Although the JAMA and GQS are widely used to evaluate online health information, they were originally designed for written materials rather than video content.[10] Therefore, they may not fully capture the audiovisual quality, clarity, or patient-centered aspects of YouTube videos.[10]

In addition, the guidelines of the American Academy of Orthopaedic Surgeons (AAOS),[17] the framework of "What information should a person searching the internet about a disease obtain?", and the "Clubfoot-Specific Score" (CSS) evaluated by two orthopedic surgeons were used.

This score aims to assess the educational quality and comprehensiveness of the videos related to pes equinovarus. Scoring systems based on the online AAOS guidelines, similar to those applied in our study, have also been used in the literature for various other orthopedic conditions (Table 1).[7,11] CSS scores ranged from 0 to 20 and were classified as excellent (16–20), good (11–15), fair (6–10), and poor (0–5). Finally, the DISCERN tool, developed in Oxford, UK, was applied. This 16-question instrument, scored from 1 to 5/item, yields a total score of 16–80, classified as very poor (16–28), poor (29–41), fair (42–54), good (55–67), and excellent (68–80).

Table 1. Clubfoot specific score

Definition	• (2 points) Defines the disease. (e.g., inward rotation of the foot, equinus position) • (1 point) Describes the prevalence of the disease. (e.g., incidence, male/female ratio) • (1 point) Discusses the etiology. (e.g., unknown exact cause, genetic and environmental factors)
Pathoanatomy	• (1 point) Describes the bone pathologies • (1 point) Describes the soft tissue pathologies. (e.g., small foot, short/thin calf muscles)
Diagnosis	• (1 point) Describes prenatal (intrauterine) diagnosis • (1 point) Discusses the distinction between idiopathic and non-idiopathic PEV.
Treatment	• (1 point) Answers the question: When should treatment be initiated? • (2 points) Provides information on the Ponseti method and other techniques (e.g., French method) • (3 points) Describes the Ponseti method (e.g., duration and number of casts, information about Achilles tenotomy) • (2 points) Provides information about orthoses (e.g., duration of use and im-portance of proper application) • (2 points) Discusses surgical treatment and procedures (e.g., management in case of relapse, types of interventions).
Prognosis and Outcome	• (1 point) Describes the consequences if the condition is left untreated • (1 point) States that treatment outcomes are favorable in idiopathic PEV.

PEV: Pes equinovarus.

Video selection was performed by authors who did not participate as observers. Links to these videos were compiled into a table and shared with the observers. Before evaluation, the observers received training on the scoring systems. Video assessment and scoring according to JAMA, GQS, DISCERN, and CSS were independently performed in a blinded manner by two pediatric orthopedic surgeons fluent in both Turkish and English. To assess intra-observer reliability, videos were re-evaluated and re-scored 1 month later by the same observers, blinded to their previous scores. Scores of Turkish and English videos were compared statistically.

Statistical Analysis

Data were analyzed using IBM Statistical Package for the Social Sciences (SPSS) Statistics version 27.0 (SPSS Inc., Chicago, IL). Continuous variables were reported as means with standard deviations, and categorical variables as absolute and relative frequencies. Relationships between usefulness scores and technical features were examined using Pearson correlation. Inter-rater agreement for video quality was assessed with the intraclass correlation coefficient (ICC) and 95% confidence intervals. ICC values <0.5 indicated poor reliability, 0.5–0.75 moderate, 0.75–0.90 good, and >0.90 excellent reliability. The mean video evaluation scores between the two groups were analyzed using Student's t-test. Statistical significance was set at $p<0.05$.

RESULTS

The duration of videos, number of views, number of days online, number of likes, and the statistical comparison between the two groups are presented in Table 2.

The intraobserver and interobserver agreements for all scoring systems were analyzed statistically. The results are presented in Table 3.

When intraobserver agreement according to the scoring systems was analyzed, a moderate level of agreement was found for the JAMA score, moderate for the GQS score, excellent for the DISCERN score, and good for the CSS.

When interobserver agreement was evaluated, a moderate level of agreement was found for the JAMA and GQS scores, excellent for the DISCERN score, and good for the CSS.

Table 2. Descriptive statistics of the videos

	Turkish (mean±SD*)	English (mean±SD*)	p**
Video duration (second)	331.04±782	350.6±340.3	=0.428
Number of views	5249±4261.7	87151±180.1	=0.003
Number of days since publication	1285.4±747.9	2204±1193.3	=0.0125
Number of likes	34.3±45.9	644.5±1157.7	<0.001

SD: Standard deviation. *: Mean±standard deviation; **: Student t-test.

Correlation analysis among the scoring systems revealed a strong, statistically significant positive association between CSS and DISCERN (r=0.970, p<0.001), a strong, statistically significant positive association between CSS and GQS (r=0.695, p<0.001), and a moderate, statistically significant positive association between CSS and JAMA (r=0.410, p=0.032).

In Turkish videos, 76.4% scored ≤2 according to the JAMA criteria, 38.3% scored ≤2 according to the GQS criteria, 55.9% were rated as poor or very poor according to DISCERN, and 54.2% were classified as poor according to CSS.

In English videos, 83.07% scored ≤2 according to the JAMA criteria, 48.8% scored ≤2 according to the GQS criteria, 58.4% were rated as poor or very poor according to DISCERN, and 63.07% were classified as poor according to CSS.

The numerical values and proportions of video sources/uploaders in the two groups are summarized in Figure 1. In the Turkish group, physician uploaders were observed at a higher proportion (p=0.04), whereas in the English group, videos uploaded by medical research groups and patients/family members were observed at a higher proportion (p=0.04).

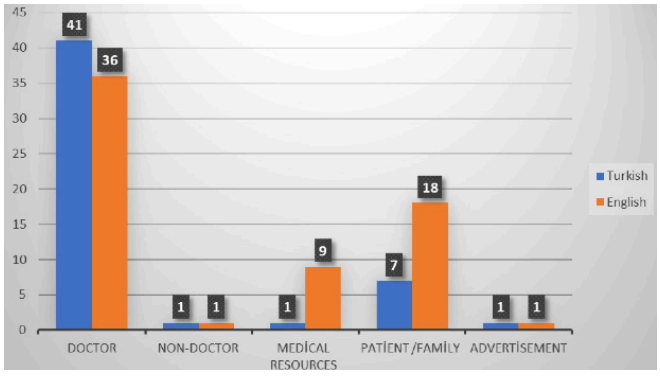

Figure 1. Comparison of two groups according to video uploaders.

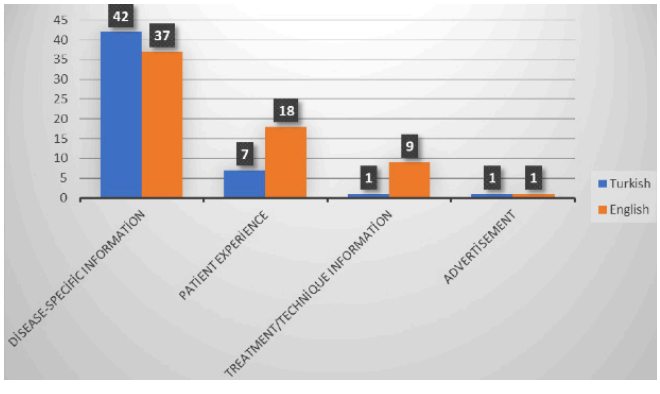

Figure 2. Comparison of videos based on content type.

When the types of content were analyzed, the numerical values and proportions for both groups are presented in Figure 2. In the Turkish videos, disease-specific information videos were observed at a higher proportion (p=0.184), whereas in the English videos, compared with the Turkish group, videos focusing on patient experience and treatment/treatment techniques were observed at a higher proportion (p=0.184).

The values of the videos according to all scoring systems and their statistical comparisons are presented in Table 4.

The scores of physician-uploaded videos and their statistical comparisons are presented in Table 5.

Table 3. Intraobserver and interobserver reliability

	Observer 1 Intraob-server ICC* (p**)	Observer 2 Intraob-server ICC* (p**)	Interobserver ICC* (p**)
JAMA	0.729 (<0.001)	0.707 (<0.001)	0.555 (<0.001)
GQS	0.646 (<0.001)	0.586 (<0.001)	0.591 (<0.001)
DISCERN	0.962 (<0.001)	0.969 (<0.001)	0.981 (<0.001)
CSS	0.844 (<0.001)	0.767 (<0.001)	0.864 (<0.001)

ICC: Intraclass correlation coefficient; JAMA: Journal of the American Medical Association; GQS: Global quality score; CSS: Clubfoot-specific score. *: Interclass correlation coefficient; **: Reliability analysis.

Table 4. Comparison of videos according to scores

	Turkish (mean±SD)*	English (mean±SD)*	p**
JAMA	2.01±0.73	1.86±0.76	=0.264
GQS	3.13±1.21	2.8±1.16	=0.171
DISCERN	41.1±22.9	23.6±17.1	<0.001
CSS	8.8±5.17	5.15±4.04	<0.001

SD: Standard deviation; JAMA: Journal of the American Medical Association; GQS: Global quality score; CSS: Clubfoot-specific score. *: Mean±standard deviation; **: Student t-test.

Table 5. Comparison of scores of videos uploaded exclusively by physicians

	Turkish (mean±SD)*	English (mean±SD)*	p*
JAMA	2.09±0.73	1.9±0.8	=0.392
GQS	3.34±1.15	2.97±1.13	=0.162
DISCERN	47.34±20.8	26.8±17.2	<0.001
CSS	10.07±4.1	6±4.1	<0.001

SD: Standard deviation; JAMA: Journal of the American Medical Association; GQS: Global quality score; CSS: Clubfoot-specific score. *: Mean±standard deviation; **: Student t-test.

DISCUSSION

The main findings of our study are that, although Turkish videos had lower views, likes, and online duration compared with English videos, they demonstrated similar reliability and educational quality in terms of JAMA and GQS scores, while DISCERN and CSS scores were higher in Turkish videos than in English videos. In recent years, social media has become one of the primary sources for patients seeking health-related information.[18,19] In this context, YouTube, with more than two billion active users, continues to be a leading online video platform.[20] However, the absence of peer review raises concerns regarding the reliability and quality of videos available on the platform.[6,21] Previous studies have evaluated the quality of videos related to upper extremity, shoulder, knee, hip, and spine conditions,[7,11,12,14,22] while there is a limited number of analyses regarding videos on pes equinovarus.[23] Our study represents the first to compare Turkish and English content on this topic.

Social media is an important tool for health communication for both institutions and individuals.[24] The online behaviors of content creators and users are influenced by the sociocultural characteristics of their communities.[25] While healthcare professionals primarily use social media for professional development, patient education, and occasionally for commercial purposes, individuals mainly use it to seek health information, share personal experiences, and participate in online support groups.[18] In our study, the higher proportion of physician-generated content in Turkish videos, in contrast to the predominance of patient-experience-based content in English videos, may reflect these sociocultural differences. Furthermore, in Türkiye, medical content production is largely physician-centered, and the online sharing of personal health experiences is relatively limited compared with Western countries. This may contribute to the presence of more structured and systematically presented content in Turkish videos, particularly regarding diagnostic and therapeutic processes.

Approximately 22–48% of health-related YouTube videos are reported to be uploaded by physicians.[26] In our study, 80% of Turkish videos and 55% of English videos were produced by physicians. The literature indicates that physician-generated content generally demonstrates higher quality.[26,27] Our findings support this observation, as physician-produced videos in both groups achieved higher quality scores. The higher contribution of physicians to Turkish videos may explain why this group received higher quality scores compared with English videos.

When only physician-uploaded videos were analyzed, Turkish content continued to demonstrate higher quality compared with English videos. This can be interpreted as follows: Videos focusing on a specific aspect or a single treatment method for pes equinovarus may score lower in standardized scoring systems. This does not imply that these videos provide insufficient information on the discussed topic.[11] However, emphasizing a single topic or specialized information may cause confusion for patients. Therefore, there is a need for comprehensive educational videos that clearly convey the full process from diagnosis to treatment options.

In most studies of this type, DISCERN and JAMA scores are widely used evaluation systems.[13-16] However, it is known that these scales were not originally designed to directly assess video sources (DISCERN) or patient education materials (JAMA).[10,13] To objectively evaluate pes equinovarus-specific content, the CSS was developed for this study. Although CSS was applied for the 1st time in this study, high intraobserver and interobserver agreement and strong positive correlations with other scoring systems support its applicability and reliability.

This study is among the first to directly compare English and Turkish YouTube content related to pes equinovarus. Our findings suggest that language differences may influence not only terminology but also content comprehensiveness, educational value, and the characteristics of video uploaders. In this respect, the study offers a new perspective to the literature by emphasizing the linguistic and cultural dimensions of online orthopedic information sources.

Our study has several limitations. First, YouTube is a constantly evolving platform that presents personalized content through artificial intelligence algorithms.[28] To minimize these effects, searches were conducted using a browser with cleared history and cookies, and only results from a specific time frame were analyzed. Second, only the first 50 videos for each search term were evaluated. Although this represents only a small portion of YouTube content, this approach has been widely used in previous studies.[10-12] Third, some videos may focus on highly specific aspects of the condition. While such videos provide detailed information on specific topics, they may have been rated as inadequate in our study due to insufficient coverage of general information. This study is that the CSS was used in this form for the 1st time. Nevertheless, similar scoring systems have been utilized in previous studies evaluating YouTube video quality for various orthopedic pathologies. Therefore, this methodological limitation can be considered acceptable and consistent with approaches reported in the existing literature.

CONCLUSION

YouTube remains an important and widely used platform for patient education and information; however, existing content generally lacks standardized and comprehensive educational materials. Our study suggests that Turkish videos demonstrate

higher quality than English videos, largely due to a greater proportion of physician-generated content. Furthermore, the CSS, developed specifically for pes equinovarus and applied in this study, appears to be a practical and reliable tool for evaluating content quality.

DECLARATIONS

Ethics Committee Approval: The study used only publicly available online data (YouTube videos) and did not include personal or identifiable information; thus, institutional review board approval was not applicable.

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: No artificial intelligence (AI) tools were used in the conduct of this research.

Authorship Contributions: Concept – TOB; Design – TOB, NE; Supervision – HG; Fundings – OS, SA; Materials – OS, SA; Data collection &/or processing – OS, SA; Analysis and/or interpretation – NE, TOB, MSS; Literature search – MSS; Writing – TOB, MSS; Critical review – NE, MSS.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Smythe T, Rotenberg S, Lavy C. The global birth prevalence of clubfoot: a systematic review and meta-analysis. EClinicalMedicine 2023;63:102178. Erratum in: EClinicalMedicine 2025;80:103087.
2. Parker SE, Mai CT, Strickland MJ, Olney RS, Rickard R, Marengo L, et al. Multistate study of the epidemiology of clubfoot. Birth Defects Res A Clin Mol Teratol 2009;85:897–904.
3. Mahan ST, Miller PE, May CJ, Kasser JR. Prospective evaluation of parental anxiety related to newborn foot disorder. J Child Orthop 2019;13:500–7.
4. Oprescu F, Campo S, Lowe J, Andsager J, Morcuende JA. Online information exchanges for parents of children with a rare health condition: key findings from an online support community. J Med Internet Res 2013;15:e16.
5. Desai T, Shariff A, Dhingra V, Minhas D, Eure M, Kats M. Is content really king? An objective analysis of the public's response to medical videos on YouTube. PLoS One 2013;8:e82469.
6. Javidan A, Nelms MW, Li A, Lee Y, Zhou F, Kayssi A, et al. Evaluating YouTube as a source of education for patients undergoing surgery: a systematic review. Ann Surg 2023;278:e712–8.
7. Yüce A, Erkurt N, İğde N, Mısır A. The quality of YouTube videos related to patellofemoral pain syndrome is insufficient. J Back Musculoskelet Rehabil 2023;36:903–10.
8. DuPont-Reyes MJ, Villatoro AP, Tang L. Health information scanning and seeking in diverse language, cultural and technological media among latinx adolescents: cross-sectional study. J Med Internet Res 2025;27:e64672.
9. Ittefaq M, Seo H, Abwao M, Baines A. Social media use for health, cultural characteristics, and demographics: A survey of Pakistani millennials. Digit Health 2022;8:20552076221089454.
10. Eysenbach G, Köhler C. How do consumers search for and appraise health information on the world wide web? Qualitative study using focus groups, usability tests, and in-depth interviews. BMJ 2002;324:573–7.
11. Yüce A, Ocak O, Gur V, Misir A. YouTube videos about plantar fasciitis and calcaneal spurs are insufficient and poor quality. J Am Podiatr Med Assoc 2023;113:22–187.
12. Erdem MN, Karaca S. Evaluating the accuracy and quality of the information in kyphosis videos shared on YouTube. Spine (Phila Pa 1976) 2018;43:E1334–9.
13. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware. JAMA 1997;277:1244–5.
14. Yuce A, Oto O, Vural A, Misir A. YouTube provides low-quality videos about talus osteochondral lesions and their arthroscopic treatment. Foot Ankle Surg 2023;29:441–5.
15. Mohammadzadeh Azarabadi J, Gökkuş K, Shalalfa N. Evaluating the quality of osteonecrosis and avascular necrosis content on youtube: a multidimensional analysis using DISCERN, JAMA, and GQS tools. J Orthop Surg Res 2025;20:695.
16. Yüce A, İğde N, Ergün T, Mısır A. YouTube provides insufficient information on patellofemoral instability. Acta Orthop Traumatol Turc 2022;56:306–10.
17. OrthoInfo. Clubfoot. Available at: https://orthoinfo.aaos.org/en/diseases--conditions/clubfoot/ Accessed Jan 19, 2026.
18. Chen J, Wang Y. Social media use for health purposes: systematic review. J Med Internet Res 2021;23:e17917.
19. Hanna G, Batko BD, Potter J, Ippolito J, Edobor-Osula F. The role of social media in clubfoot: information sharing and social support. J Child Orthop 2021;15:81–8.
20. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. BMC Med Educ 2022;22:382.
21. Mohamed F, Shoufan A. Users' experience with health-related content on YouTube: an exploratory study. BMC Public Health 2024;24:86.

22. Kunze KN, Alter KH, Cohn MR, Vadhera AS, Verma NN, Yanke AB, et al. YouTube videos provide low-quality educational content about rotator cuff disease. Clin Shoulder Elb 2022;25:217–23.
23. Ranade AS, Belthur MV, Oka GA, Malone JD. YouTube as an information source for clubfoot: a quality analysis of video content. J Pediatr Orthop B 2020;29:375–8.
24. Patrick M, Venkatesh RD, Stukus DR. Social media and its impact on health care. Ann Allergy Asthma Immunol 2022;128:139–45.
25. Khoong EC, Le GM, Hoskote M, Rivadeneira NA, Hiatt RA, Sarkar U. Health information-seeking behaviors and preferences of a diverse, multilingual urban cohort. Med Care 2019;57(Suppl 6 2):S176–83.
26. Sari E, Umur LF. Quality analysis of hallux valgus videos on YouTube. J Am Podiatr Med Assoc 2021;111.
27. Jildeh TR, Abbas MJ, Abbas L, Washington KJ, Okoroha KR. YouTube Is a poor-quality source for patient information on rehabilitation and return to sports after hip arthroscopy. Arthrosc Sports Med Rehabil 2021;3:e1055–63.
28. Yesilada M, Lewandowsky S. Systematic review: YouTube recommendations and problematic content. Internet Policy Rev 2022;11:1652.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2026.94547
Eur Arch Med Res 2026;42(2):169–173

Differentiating Physical and Mental Fatigue in Pregnancy: A Chalder-Based Quantitative Analysis

Elif Ucar, Melih Bestel

Department of Midwifery, Istanbul Esenyurt University, Faculty of Health Sciences, Istanbul, Türkiye

ABSTRACT

Objective: The aim of this study was to compare the levels of physical and mental fatigue during pregnancy and to determine which type of fatigue is dominant. In addition, the relationships between fatigue levels and some demographic and hematological variables were also investigated.

Materials and Methods: In this cross-sectional study, 222 pregnant women admitted to the obstetrics outpatient clinic of a tertiary health institution were included. Physical and mental fatigue levels were assessed with the sub-dimensions of the Chalder Fatigue Questionnaire, and subjective fatigue perception was measured with the Visual Analog Scale (VAS). In addition, participants' clinical data such as age, gravida, parity, hemoglobin, iron and ferritin levels were analysed.

Results: The Chalder physical fatigue subscore was significantly higher than the mental fatigue subscore ($p<0.001$). A significant positive correlation was found between VAS scores and CHALDER total scores ($\rho=0.469$, $p<0.001$). No statistically significant correlation was found between fatigue levels and age, gravida, parity, hemoglobin, iron and ferritin levels ($p>0.05$).

Conclusion: The most prominent type of fatigue in pregnancy is physical fatigue. The level of fatigue may be related to pregnancy-specific physiological and psychosocial stress factors rather than hematological or demographic factors. It is recommended that a holistic and multidisciplinary approach should be adopted in the management of fatigue in pregnancy.

Keywords: Chalder, Fatigue, Mental fatigue, Physical fatigue, Pregnancy, Visual analog scale

Cite this article as: Ucar E, Bestel M. Differentiating Physical and Mental Fatigue in Pregnancy: A Chalder-Based Quantitative Analysis. Eur Arch Med Res 2026;42(2):169–173.

INTRODUCTION

Fatigue during pregnancy is a condition that women frequently experience throughout their pregnancy. This condition is influenced by various physiological, psychological and social factors. In the first trimester of pregnancy, all the attention of the pregnant woman is on herself. Hormonal changes, metabolic increase and physiological loads in the body during pregnancy may increase the feeling of fatigue.[1] Mood changes may also trigger a mother's fatigue during pregnancy. Studies have shown that psychological conditions such as stress, anxiety and depression are frequently observed during pregnancy, and these conditions may increase the feeling of fatigue.[2]

In addition, an increase in body mass and changes in physical activity level during pregnancy are also important factors affecting fatigue. Pregnant women may often have to spend

Address for correspondence: Elif Ucar. Department of Midwifery, Istanbul Esenyurt University, Faculty of Health Sciences, Istanbul, Türkiye
E-mail: eliflyy@hotmail.com **ORCID ID:** 0000-0001-5302-4688

Submitted: 29.08.2025 **Revised:** 24.12.2025 **Accepted:** 05.01.2026 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

more energy to perform their daily activities. This may increase physical fatigue and affect the general quality of life.[1] Especially in the prenatal period, restlessness, insomnia and difficulties encountered in daily activities may increase physical fatigue.[3] In addition, it has been emphasised that fatigue experienced during pregnancy may continue in the postpartum period.[4]

Studies show that approximately 20% of pregnant women experience symptoms of depression during pregnancy, and this may increase mental fatigue.[5] In addition, psychological problems experienced during pregnancy lead to mental fatigue by making psychological adaptation of many women difficult.[6] One of the main causes of mental fatigue is inadequate social support. Active participation of spouses or family members in the pregnancy process may positively affect the psychological well-being of the mother. A positive social support network may minimise mental fatigue by reducing the stress levels of pregnant women.[7] However, negative communication and conflicts experienced with the partner during pregnancy may increase anxiety and mental fatigue.[6] In addition, physical changes experienced during pregnancy may also contribute to mental fatigue. Hormonal changes and physical disorders may lead to mood fluctuations and affect the mind-body connection.[8]

In this study, it was aimed to determine the dominant type of fatigue occurring during pregnancy.

MATERIALS AND METHODS

In this retrospective study, 222 pregnant women who applied to the pregnancy outpatient clinic of a 3rd level hospital were included. The study protocol was approved by the Istanbul Esenyurt University Ethics Committee (Approval No: 2025/06, Date: 05.08.2025), and the study was designed in accordance with the Declaration of Helsinki.

Healthy pregnant patients were included in the study. Patients with any muscle disease, using psychological drugs, diagnosed with fibromyalgia before pregnancy, gestational diabetes or any type of diabetes, diagnosed with hypertension or preeclampsia, heart disease, or non-pregnant patients were not included.

Demographic data and blood analyses during pregnancy were recorded, and fatigue levels of the patients were evaluated with the Chalder fatigue scale.

The Chalder fatigue scale is a self-report scale with high psychometric validity developed to assess the fatigue levels of individuals. There are two subcategories in this scale consisting of 11 items in total. Seven questions assess physical fatigue, and four questions assess mental fatigue. Each item is evaluated according to the level of fatigue experienced in the last 2 weeks.[9]

Statistical Analysis

The data obtained in this study were analysed using the Statistical Package for the Social Sciences v.26.0 software. Descriptive statistics for numerical and categorical variables were calculated and presented in tables. Mean±standard deviation (SD), median, minimum and maximum values were calculated for continuous variables. Frequency (*n*) and percentage (%) distributions were given for categorical variables. The assumption of normal distribution was evaluated by Shapiro–Wilk test. When Shapiro–Wilk $p<0.05$, it was accepted that the variable was not normally distributed.

RESULTS

The mean age of 222 women included in the study was 28.12±5.84, and demographic data are given in Table 1.

Only 1 of the pregnant women included in the study used folic acid before pregnancy, and the rest of the patients did not take folic acid, iron supplements or B12 supplements before pregnancy. Biochemical values of the pregnant women are given in Table 2.

Table 1. Distribution of general characteristics of the study

Variable (*n*=222)	Average±SD	Median (Min–Max)
Age	28.12±5.84	27.0 (17.0–43.0)
Gestation week	14.74±8.63	12.0 (4.0–38.0)
Gravida	3.42±2.19	3.0 (1.0–13.0)
Parity	1.98±1.87	2.0 (0.0–9.0)
Abort	0.43±0.97	0.0 (0.0–8.0)

SD: Standard deviation; Min: Minimum; Max: Maximum.

Table 2. Distribution of biochemical characteristics

Variable	Average±SD	Median (Min–Max)
B12 Level	236.22±125.95	212.5 (83.0–1478.0)
Folic acid level	8.99±4.04	8.05 (2.9–20.0)
Hemoglobin	12.36±1.16	12.5 (8.7–15.5)
Hematocrit	36.88±2.98	37.15 (28.4–45.86)
MCV	82.38±7.09	83.4 (38.6–95.1)
Iron	80.04±44.87	78.0 (9.0–405.0)
Iron binding	276.35±96.82	266.0 (19.0–593.0)
Ferritin	25.94±21.48	19.47 (1.0–110.31)

SD: Standard deviation; Min: Minimum; Max: Maximum; MCV: Mean corpuscular volume.

Table 3. Distribution of Chalder and visual analog scale scores

Variable	Average±SD	Median (Min–Max)
Chalder physical fatigue score	8.15±2.43	7.0 (7.0–21.0)
Chalder mental fatigue score	4.01±0.1	4.0 (4.0–5.0)
Chalder total points	12.15±2.44	11.0 (11.0–25.0)
Visual analogue scale	2.14±1.74	1.0 (1.0–8.0)

SD: Standard deviation; Min: Minimum; Max: Maximum.

Fatigue scales of the patients are given in Table 3.

In the study, a strong positive correlation was observed between Chalder fatigue scores and visual analogue scale (VAS) scores. This indicates that the subjective fatigue perceptions of individuals are consistent with the objective fatigue levels measured by the Chalder scale. As the subjectively reported fatigue level VAS increases, the objective fatigue level measured by the Chalder questionnaire also increases. (Spearman correlation coefficient $\rho=0.71$, $p<0.0001$ in the analysis. Since normal distribution assumptions were not met (Shapiro–Wilk test $p<0.05$ for both variables) Spearman test was preferred (Fig. 1).

In the study, a strong and significant positive correlation was found between Chalder physical fatigue subscale and VAS scores. This result shows that the physical fatigue level of the individual during pregnancy is compatible with subjectively perceived general fatigue. (Spearman correlation coefficient $\rho=0.79$, $p<0.0001$ in the analysis).

A statistically significant but weak positive correlation was found between Chalder mental fatigue score and VAS. This result suggests that the general fatigue perception of individuals is associated with physical fatigue rather than mental fatigue. (Spearman correlation coefficient $\rho=0.177$, $p=0.000$ in the analysis).

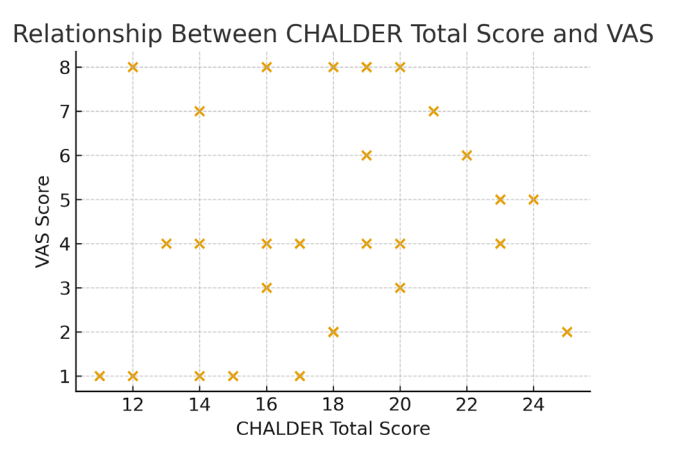

Figure 1. Relationship between Chalder total score and visual analogue scale

Table 4. Comparison of Chalder total fatigue scores according to pregnancy trimester

Gestational trimester	Mean±SD	Median (Min–Max)	*n*
1st Trimester (1–14 weeks)	12.01±2.51	11.0 (11.0–25.0)	126
2nd trimester (15–28 weeks)	12.0±2.18	11.0 (11.0–19.0)	60
3rd trimester (29–42 weeks)	13.26±2.64	13.0 (11.0–18.0)	36

SD: Standard deviation; Min: Minimum; Max: Maximum.

There was a statistically significant difference between Chalder total scores according to gestational week ($p<0.01$). The highest fatigue level was observed in the 3rd trimester (29–42 weeks) group. This finding indicates that physical and mental fatigue increase as the pregnancy progresses (Table 4).

DISCUSSION

In this study, the Chalder Fatigue Scale, which is used to assess physical and mental fatigue in pregnant women, was compared with the VAS, which measures the subjective perception of fatigue, and the relationship between the two different fatigue measurement methods was analysed. The results showed that as the subjective fatigue level reported by VAS increased, the CHALDER total score also increased significantly. This result supports the usefulness of VAS for rapid screening in the clinic, but suggests that comprehensive scales such as CHALDER remain important for multidimensional fatigue assessment.

Pregnancy is a multidimensional and sensitive period in which many physiological and psychological changes occur simultaneously in the female body. Hormonal fluctuations, hematological changes, increased metabolic needs and compulsory changes in lifestyle facilitate the emergence of fatigue. In the first trimester of pregnancy, women's efforts to adapt to the rapid changes in their bodies, combined with symptoms such as a tendency to sleep, nausea, vomiting and general weakness, lead to an increase in both physical and mental fatigue levels. Indeed, in the Generation R study conducted by Bai et al.,[10] it was shown that symptoms such as nausea, vomiting and fatigue observed in early pregnancy had negative effects on quality of life. However, Chou et al.[11] emphasised that fatigue, nausea and vomiting experienced in early pregnancy were closely related with depressive symptoms. This shows that pregnancy fatigue has not only biological but also psychological components. In this study, we did not find a significant relationship between hematological

data and fatigue, which shows that we obtained similar results with the literature.

The findings of this study also provide important clinical insight into the multidimensional structure of antenatal fatigue. The significantly higher physical fatigue scores compared with mental fatigue scores highlight the predominance of somatic burdens during pregnancy, particularly as gestation progresses. Physiological mechanisms such as increased metabolic demands, sleep disturbance, musculoskeletal strain, and weight gain may explain this pattern, suggesting that physical fatigue should be considered a primary focus in antenatal follow-up. Although mental fatigue was also detected, its weaker correlation with subjective fatigue perception indicates that psychological components, while relevant, may be secondary to the physical stressors experienced by pregnant women. These findings support the importance of routine assessment of physical fatigue symptoms during prenatal care, particularly in the third trimester, where fatigue appears to peak.

In this study, the relationship between Chalder total scores and age, gravida, parity, hemoglobin, iron, and ferritin levels was examined, and none of these parameters showed a statistically significant relationship with fatigue score ($p>0.05$). This finding suggests that the main determinants of fatigue in pregnancy are not hematological but rather psychosocial and behavioral factors. However, the weak correlation between the Chalder mental fatigue score and the VAS suggests that the physical component of fatigue may also be important. The fact that fatigue increased in the third trimester of pregnancy in our analysis also supports this. In the study conducted by Zhang et al.[12] in the third trimester, it was shown that pregnancy fatigue was significantly associated with factors including depression, sleep quality, and cognitive dysfunction. Similarly, in a study conducted by Wang et al.[13] using the latent class analysis method in 2022, pregnancy fatigue was classified in different profiles, and it was found that variables including advanced gestational week, adverse obstetric history, and poor sleep quality significantly affected the level of fatigue. In addition, it has been reported that prenatal fatigue is not limited to the gestational period and may also be related with depressive symptoms that may develop in the postpartum period. Fairbrother et al.[14] showed that individuals with high levels of fatigue during pregnancy had a higher risk of developing major depression in the postnatal period. Therefore, pregnancy fatigue should be evaluated not only as a symptom but also as a predictor of postpartum psychological health.

The absence of a significant association between fatigue severity and hematological or demographic parameters underscores an important conceptual shift. Routine laboratory markers frequently used in antenatal practice may not adequately capture the complexity of pregnancy-related fatigue. Instead, psychosocial, behavioral, and physiological contributors are likely to play a greater role in shaping fatigue perception. This interpretation aligns with existing evidence, which emphasizes that sleep quality, emotional state, stress level, and social support exert stronger effects on fatigue than biological indicators alone. Accordingly, integrating psychosocial screening tools into routine antenatal assessments – alongside physical evaluation – may strengthen early identification and management strategies for fatigue in pregnancy.

One of the key strengths of this study is its relatively large sample size of 222 pregnant women, which enhances the statistical reliability and generalizability of the findings. In addition, the study simultaneously evaluated both physical and mental components of fatigue, allowing for a more comprehensive understanding of the multidimensional nature of fatigue during pregnancy. Another methodological strength lies in the combined use of the Chalder Fatigue Scale and the VAS, enabling the integration of both objective multidimensional fatigue scoring and subjective fatigue perception. This dual-scale approach strengthens the methodological robustness of the study and provides a more nuanced interpretation of fatigue patterns among pregnant women.

Limitation

Limitations of the study: Although the data of 222 pregnant women in this study are considered sufficient for statistical analyses, studies with larger samples are needed. Since the study had a cross-sectional design, the findings obtained belonged only to a certain period of time, and it was not possible to evaluate the change in fatigue levels during pregnancy over time. Psychosocial variables such as depression, anxiety, social support, sleep quality and stress level, which may affect the level of fatigue during pregnancy, were not systematically evaluated in this study. The effects of these variables on fatigue could not be taken into account. The study was conducted with a group of patients from a single health institution, and the generalisability of the results to different socioeconomic and cultural levels is limited.

CONCLUSION

When evaluating fatigue during pregnancy, it is important to focus not only on hematological or demographic indicators but also on parameters such as psychosocial status, sleep patterns, emotional burden, and level of social support. However, it should not be forgotten that mental fatigue may be accompanied by a physical component, especially in third-trimester pregnancies. Adopting such a holistic approach in clinical practice will both improve maternal health and contribute to the prevention of mental health problems that may develop in the postnatal period.

DECLARATIONS

Ethics Committee Approval: The study protocol was approved by the Istanbul Esenyurt University Ethics Committee (Approval No: 2025/06, Date: 05.08.2025).

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: This article was produced without the use of artificial intelligence(AI)- assisted technologies , including Large Language Models (LLMs) chatbots, or image generators.

Authorship Contributions: Concept – EU; Design – EU, MB; Supervision – MB; Fundings – EU; Materials – EU; Data collection &/or processing – EU, MB; Analysis and/or interpretation – MB; Literature search – EU; Writing – EU, MB; Critical review – MB.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Sarıyıldız A, Benlidayı, İC. Gebelik dönemindeki potansiyel kas-iskelet sistemi sorunları. akdt 2022;31:279–83. [Article in Turkish]
2. Demir A, Yazar, M. Prenatal dönemde stres, anksiyete ve depresyonun incelenmesi: geleneksel derleme. KTOKUSB-D 2024;4:97–211. [Article in Turkish]
3. Sabancı Baransel E, Uçar T. Riskli gebelerde fiziksel aktivite, gebelikte iyilik ve gebelikteki yakınmaların yaşam kalitesine etkisi: karşılaştırmalı bir çalışma. Anatolian J Health Res 2021;2:101–6. [Article in Turkish]
4. Özdamar Ö, Yılmaz O, Beyca HH, Muhçu M. Gebelik ve postpartum dönemde sık görülen ruhsal bozukluklar. Zeynep Kamil Tıp Bülteni 2014;45:71–7. [Article in Turkish]
5. Yılmaz M, Yar D. Gebelik ve postpartum dönemde kadın ruh sağlığı: derleme çalışması. Adnan Menderes Üniversitesi Sağlık Bilimleri Fakültesi Dergisi 2021;5:93–100. [Article in Turkish]
6. Güner Ö. Gebelikte eş desteğinin stres, anksiyete ve depresyonla ilişkisinin incelenmesi. EGEHFD 2023;39:235–42. [Article in Turkish]
7. Özcan H, Arar İ, Çakır A. Babalar ve gebelik süreci. Zeynep Kamil Tıp Bülteni 2018;49:72–6. [Article in Turkish]
8. Kızılkaya Beji N, Murat M, Köse S. Perinatal dönem ruh sağlığı sorunları ve hemşirelik yaklaşımı. Black Sea J Health Sci 2022;5:116–23. [Article in Turkish]
9. Chalder T, Berelowitz G, Pawlikowska T, Watts L, Wessely S, Wright D, et al. Development of a fatigue scale. J Psychosom Res 1993;37:147–53.
10. Bai G, Korfage IJ, Groen EH, Jaddoe VW, Mautner E, Raat H. Associations between nausea, vomiting, fatigue and health-related quality of life of women in early pregnancy: the generation R study. PLoS One 2016;11:e0166133.
11. Chou FH, Lin LL, Cooney AT, Walker LO, Riggs MW. Psychosocial factors related to nausea, vomiting, and fatigue in early pregnancy. J Nurs Scholarsh 2003;35:119–25.
12. Zhang X, Cao D, Sun J, Shao D, Sun Y, Cao F. Sleep heterogeneity in the third trimester of pregnancy: Correlations with depression, memory impairment, and fatigue. Psychiatry Res 2021;303:114075.
13. Wang RY, Xıng W, FA XE, Chen JZ. Latent class analysis of fatigue of pregnant women in late pregnancy. J Nurs 2022;29:11–6.
14. Fairbrother N, Hutton EK, Stoll K, Hall W, Kluka S. Psychometric evaluation of the Multidimensional Assessment of Fatigue scale for use with pregnant and postpartum women. Psychol Assess 2008;20:150–8.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2025.75010
Eur Arch Med Res 2026;42(2):174–182

Nurses' Attitudes Toward Patients with Chronic Pain

Ozge Demir, Ezgi Demirturk Selcuk

First and Emergency Aid Program, Vocational School of Health Services, Erzincan Binali Yildirim University, Erzincan, Turkiye

ABSTRACT

Objective: Chronic pain represents a complex health issue with biopsychosocial dimensions that requires a multidisciplinary approach for effective management. Knowledge deficits and negative attitudes of nurses toward patients with chronic pain constitute significant barriers to quality pain care and adversely affect patient outcomes. This study aimed to determine the attitudes of nurses toward patients with chronic pain.

Materials and Methods: This descriptive study was conducted with 216 nurses working in a training and research hospital. Data were collected using a participant information form and the attitudes of healthcare providers toward patients with chronic pain scale. Data were analyzed using the Statistical Package for the Social Sciences for Windows, Version 22. Descriptive statistics, Student's t-test, one-way analysis of variance, Mann–Whitney U-test, and correlation analysis were used. All results were evaluated at a significance level of $p<0.05$.

Results: It was found that 57.4% of the nurses had not received training on chronic pain management, 55.1% considered themselves competent in managing chronic pain, and 65.3% had a family member with chronic pain. The nurses' overall mean score on the attitudes toward patients with chronic pain scale was 3.40±0.46. Among the subdimensions, the sensitivity subscale mean score was 3.99±0.71, and the misconceptions subscale mean score was 2.96±0.77.

Conclusion: It was observed that nurses' attitude scores toward patients with chronic pain were at a moderate level. It was determined that receiving education on chronic pain and the unit in which the nurses worked had a positive effect on the total attitude score.

Keywords: Attitude, Chronic pain, Nurses' attitude, Pain management

Cite this article as: Demir O, Demirturk Selcuk E. Nurses' Attitudes Toward Patients with Chronic Pain. Eur Arch Med Res 2026;42(2):174–182.

INTRODUCTION

Pain is an inevitable part of human life. It is a comprehensive health problem that not only affects individuals but also has societal dimensions. The International Association for the Study of Pain (IASP) defines pain as "an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage".[1] Although there is no clear threshold for when acute pain becomes chronic, it is generally accepted that pain persisting for 3 months after the expected recovery period (according to the International Classification of Diseases, 11th edition criteria) is pathological.[2] The fact that chronic pain affects more than 30% of people globally reveals the magnitude of this problem.[3]

Address for correspondence: Ozge Demir. First and Emergency Aid Program, Vocational School of Health Services, Erzincan Binali Yildirim University, Erzincan, Turkiye

E-mail: ozgeakbaba24@gmail.com **ORCID ID:** 0000-0003-0228-1558

Submitted: 05.09.2025 **Revised:** 24.09.2025 **Accepted:** 03.10.2025 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Living with chronic pain means struggling with daily challenges around simple tasks, often being disbelieved and stigmatized. This situation reaches alarming proportions in terms of poor mental health, socioeconomic disadvantage, and societal costs. The Lancet's chronic pain series refutes the common misconception that chronic pain is an indicator of ongoing injury, instead emphasizing that it is a product of abnormal neural signaling and requires a multimodal treatment approach due to biopsychosocial dimensions.[4] The fact that pain is always subjective constitutes the fundamental paradigm of the assessment process. Although many ways are considered for the presence and evaluation of pain, the patient's pain report should be accepted as it is.[3]

The presence of pain directs individuals to seek healthcare. Proper assessment, identification of pain, and planning of appropriate interventions are fundamental requirements for effective management. Nurses, as healthcare professionals who interact most with patients, assume a critical role in pain management.[5] The knowledge level and attitudes of nurses directly affect not only pain control but also the physiological and psychological well-being of patients.[6] The translation of theoretical knowledge into clinical practice, regular pain assessments, and documentation form the foundation of quality pain care.[7] In addition, the use of quality indicators related to pain is important for evaluating the effectiveness of nursing practices.[8] However, knowledge deficits and inadequate attitudes are common among nurses and constitute a significant barrier to effective pain management.[9-11] Previous research also indicates that nurses worldwide have negative attitudes toward pain assessment and management.[12,13] Despite the ethical responsibility in pain management, studies conducted in various clinical settings show that nurses often tend to underestimate patients' pain.[14-18] Studies reveal that some erroneous attitudes toward patients with pain still persist among nurses. These attitudes include the widespread belief that patients exaggerate their pain, the idea that placebo use can be effective in assessing the reality of pain, and encouraging patients to show tolerance to pain before giving analgesics.[19-21] The basis of this situation lies in the difference between the nurse's perception and the patient's self-assessment.[22] Furthermore, this problem is not limited to nursing education alone; similar knowledge deficits are observed in medical education. Particularly, inadequacies in chronic pain management make pain management, which requires a multidisciplinary approach, even more complex.[23,24]

Improving nurses' attitudes toward pain management is of critical importance for both improving individual patient outcomes and enhancing healthcare quality. IASP's declaration of 2022 as the "Global Year for Translating Pain Knowledge into Practice" emphasizes the knowledge-practice gap in this field. The fact that pain management is a nursing-sensitive indicator reveals the decisive role of nurses in this process.[25,26] Nurses in contemporary healthcare systems face an intensive clinical workload characterized by increasing patient acuity and expanding care responsibilities. The systematic analysis of this situation's impact on nursing performance and patient outcomes represents a critical imperative for healthcare delivery effectiveness. This research aims to determine nurses' attitudes toward patients with chronic pain and contribute to the literature.

MATERIALS AND METHODS

Design

The study is designed as a descriptive research.

Setting and Sample

The population of the study consisted of 490 nurses working at the Erzincan Binali Yıldırım University-affiliated Training and Research Hospital located. Power analysis was conducted using G-power 3.1 program, and the sample size was determined to be 216 nurses with a 95% confidence interval and α=0.05 significance level.

Data Collection Process

The study was conducted between January 2024 and June 2024. Data collection began by providing the participants with verbal information about the purpose of the study. Informed consent, both verbal and written, was obtained from the participants who voluntarily agreed to take part in the study. Data were collected through face-to-face interviews at the times convenient for the participants. The data collection process for each participant took approximately 10 min.

Ethical and Institutional Approvals

To conduct the study, ethical approval was obtained from the Ethics Committee of the Faculty of Health and Sports Sciences at Erzincan Binali Yıldırım University (293321-08/04) on September 06, 2023, and institutional permission was granted by the institution where the research was conducted (231578024). Before data collection, the nurses were provided with detailed information about the study, and written and verbal informed consent was obtained from them. The study adhered to the Helsinki Declaration on Human Rights throughout the research process, as individual rights must be protected.

Data Collection Instruments

The research employed two primary data collection tools administered through direct face-to-face interviews: A demographic questionnaire and the Healthcare Professionals' Attitudes toward Patients.

Table 1. Participant information form.

Variable	*n*	%
Sex		
Female	169	78.2
Male	47	21.8
Education		
High school	10	4.6
Associate degree	39	18.1
Bachelor's degree	145	67.1
Graduate degree	22	10.2
Marital status		
Married	110	50.9
Single	106	49.1
Worked department		
Inpatient services	69	31.9
Surgical services	53	24.5
Intensive care unit	52	24.1
Emergency department	42	19.4
Work experience		
<1 year	13	6.0
1–5 year	94	43.5
6–10 year	41	19.0
11 years and older	68	31.5
Chronic pain management training		
Yes	92	42.6
No	124	57.4
Perceived competence in chronic pain management		
Yes	119	55.1
No	97	44.9
Chronic pain in family members		
Yes	75	34.7
No	141	65.3
Age	Mean±SD: 30.68±6.790	(Minimum: 21, Maximum: 53)

SD: Standard deviation

Participant Information Form

This form consisted of nine items assessing participant characteristics such as age, gender, relationship status, level of education, working unit, duration of professional experience, history of education in chronic pain management, self-perceived competence in managing chronic pain, and the presence of a family history of chronic pain.

Healthcare Professionals' Attitudes toward Patients with Chronic Pain Scale (HCPAPCP)

The study utilized a psychometric instrument developed by Sucu Çakmak et al., [27] which underwent comprehensive psychometric validation procedures. This tool demonstrates established validity and reliability for measuring healthcare providers' perspectives regarding individuals experiencing persistent pain conditions.

The scale's psychometric properties were confirmed through multiple analytical approaches. Reliability assessment incorporated internal consistency measures and test-retest stability evaluations. Construct validity was established using both exploratory and confirmatory factor analytical procedures. The final instrument comprises 18 items measured on a five-point Likert response format.

Factor analysis identified two distinct dimensions within the scale structure:

Sensitivity Orientation Dimension

This subscale encompasses items reflecting compassionate, understanding, and empathetic perspectives toward chronic pain sufferers. It includes items 1, 3, 4, 6, 7, 9, 11, 12, 14, 16, and 18, demonstrating strong internal consistency (Cronbach's α=0.88) among healthcare workers.

Misconception Orientation Dimension

This subscale captures negative, judgmental, stigmatizing, and erroneous beliefs regarding chronic pain patients. It incorporates items 2, 5, 8, 10, 13, 15, and 17, with acceptable reliability (Cronbach's α=0.75) in the healthcare professional sample.

Scoring Protocol

Each dimension receives independent scoring without established threshold values. The sensitivity dimension employs standard scoring methodology, given its positive item content: 1 (Strongly disagree) through 5 (Strongly agree). Scores are aggregated and divided by eleven items, yielding a range from 1 to 5. Higher values reflect greater empathetic orientation.

The misconception dimension utilizes reverse scoring due to negative item formulation: 5 (Strongly disagree) through 1 (Strongly agree). Scores are totaled and divided by seven items, producing a 1–5 range. Elevated scores indicate fewer misconceptions.

Consequently, higher composite scores across both dimensions signify more favorable attitudes toward chronic pain patients, while lower scores suggest less positive

Table 2. HCPAPCP total and subscale mean scores.

Scale/Subscale	*n*	Minimum	Maximum	Mean±SD
Sensitivity orientation	216	1.64	5	3.99±0.71
Misconception orientation	216	1	5	2.96±0.77
HCPAPCP scale	216	1.44	4.78	3.40±0.46

SD: Standard deviation; HCPAPCP: Healthcare professionals' attitudes toward patients with chronic pain.

Table 3. HCPAPCP scale and subscale score correlations.

	1	2	3
HCPAPCP scale	-		
R			
p			
Sensitivity orientation	0.781**	-	
R	0.000		
p			
Misconception orientation	0.425**	–0.233**	-
R			
p	0.000	0.001	

**$p<0.01$. HCPAPCP: Healthcare professionals' attitudes toward patients with chronic pain

professional perspectives. The current study achieved reliability coefficients of α=0.907 for sensitivity and α=0.760 for misconception factors.[27]

Statistical Analysis

Statistical analysis was conducted using the Statistical Package for the Social Sciences version 22.0 for Windows. Distribution normality was assessed through Kolmogorov–Smirnov testing. Descriptive statistics included frequencies, percentages, means, and standard deviations.

Inferential analyses employed Student's *t*-tests and One-way analysis of variance (ANOVA) for normally distributed categorical comparisons, while Mann–Whitney U tests addressed non-parametric data. Pearson correlation analysis examined bivariate relationships. *Post hoc* comparisons following significant ANOVA results utilized Games-Howell procedures when homogeneity of variance assumptions were violated. Statistical significance was established at $p<0.05$ with 95% confidence intervals for all analyses.

RESULTS

Table 1 shows that nurses averaged 30.68±6.790 years of age, with 78.2% female, 67.1% bachelor's educated, and 50.9% married. The findings show 31.9% worked in internal medicine, 43.5% had 1–5 years of experience, 57.4% lacked chronic pain training, 55.1% considered themselves competent in chronic pain management, and 65.3% had chronic pain in their families.

Table 2 shows that the sensitivity orientation dimension of the HCPAPCP achieved a mean of 3.99±0.71, while the misconception orientation dimension scored 2.96±0.77, and the overall mean score of the attitudes toward patients with chronic pain scale as 3.40±0.46.

Table 3 shows the correlation findings, indicating a robust positive association between chronic pain patient attitudes and sensitivity orientation ($r=0.781$, $p<0.01$), alongside a moderate positive correlation with misconception orientation ($r=0.425$, $p<0.01$). Table 3 further shows a statistically significant weak inverse relationship between sensitivity and misconception orientations ($r=-0.233$, $p<0.01$).

Table 4 shows notable variations in the sensitivity orientation subscale scores, a component of HCPAPCP, gender, working unit, and chronic pain management training status. Post-hoc analysis results displayed in Table 4 revealed statistically significant differences between internal medicine staff and those employed in surgical units and intensive care departments ($p<0.05$), whereas no meaningful distinctions emerged among the remaining age categories.

Table 4 presents several significant findings regarding the HCPAPCP subscales. The misconception subscale showed notable variations based on professional experience duration and chronic pain training history. Similarly, overall HCPAPCP scores differed significantly according to workplace unit and chronic pain training background ($p<0.05$). Conversely, no statistically significant differences were observed between HCPAPCP and its components in relation to self-assessed chronic pain management competency or family history of chronic pain conditions ($p>0.05$).

DISCUSSION

The evaluation of attitudes toward pain is an important step in terms of improving nurses' clinical practices and enhancing

Table 4. Comparison of participants' HCPAPCP scale scores

	Sensitivity orientation		Misconception orientation		HCPAPCP scale	
	$\bar{X}$±SD	Test *p*	$\bar{X}$±SD	Test and *p*	$\bar{X}$±SD	Test and p
Sex						
Female	4.09±0.57	Z:1.992	2.46±0.73	t=0.100	3.45±0.35	Z:-1.782
Male	3.67±1.00	*p*:0.046	2.45±0.92	*p*=0.921	3.19±0.72	*p*:0.075
Education						
High school	4.41±0.36	F:2.412	2.34±0.90	F:0.376	3.61±0.40	F:1.45
Associate degree	3.79±0.78	*p*:0.068	0.54±0.66	*p*:0.770	3.31±0.55	*p*:0.228
Bachelor's degree	4.03±0.66		2.46±0.80		3.42±0.40	
Graduate degree	3.94±0.91		0.77±0.05		3.32±0.64	
Marital status						
Married	3.97±0.70	Z:0.697	2.49±0.77	*t* 0.564	3.40±0.47	Z:-0.389
Single	4.02±0.72	*p*:0.486	2.43±0.77	*p*:0.573	3.40±0.46	*p*:0.690
Worked department						
Inpatient services[a]	4.23±0.47	F:4.692	2.48±0.67	F:1.547	3.55±0.32	KW:10.249
Surgical services[b]	3.96±0.65	*p*:0.003	2.55±0.79	*p*:0.203	3.41±0.41	*p*:0.017
Intensive care unit[c]	3.77±0.86	a>b	2.52±0.78		3.28±0.53	
Emergency department[d]	3.93±0.81	a>c	2.24±0.86		3.27±0.56	
Work experience						
<1 year	3.97±0.72	F:0.474	2.16±0.79	F:2.678	3.27±0.46	KW:2.822
1–5 year	4.06±0.65	*p*:0.701	2.43±0.82	*p*:0.048	3.42±0.42	*p*:0.420
6–10 year	3.91±0.95		2.31±0.73		3.28±0.66	
11 years and older	3.97±0.62		2.65±0.68		3.46±0.36	
Chronic pain management training						
Yes	4.22±0.59	Z:4.259	2.58±0.79	*t*:2.624	3.47±0.36	*t*:2.208
No	3.83±0.74	*p*:0.000	2.30±0.72	*p*:0.009	3.34±0.52	*p*:0.028
Perceived competence in chronic pain management						
Yes	4.08±0.64	Z:1.633	2.43±0.80	*t*:0.624	3.44±0.40	Z:–0.872
No	3.89±0.77	*p*:0.102	2.50±0.73	*p*:0.534	3.34±0.53	*p*:0.383
Chronic pain in family members						
Yes	3.98±0.54	Z:0.855	2.54±0.70	*t*:1.141	3.42±0.33	*t*:0.667
No	4.00±0.79	*p*:0.392	2.42±0.80	*p*:0.255	3.38±0.52	*p*:0.506

$\bar{X}$: Mean, SD: Standard deviation, *t*: Independent *t*-test, Z: Man–Whitney U test, F: One-way ANOVA, KW: Kruskal–Wallis, HCPAPCP: Healthcare professionals' attitudes toward patients with chronic pain

care quality.[6] Nurses are in the position of first-line advisors and supporters for patients suffering from chronic pain. Although nurses hold a pivotal position in the management of chronic pain, current evidence indicates that pain continues to be undertreated and remains a persistent global health concern.[9,28,29] Therefore, there is a need to examine attitudes toward patients with chronic pain and to understand the determinants of attitudes in further studies.[30] In the literature, there are studies on general pain and acute pain attitudes, but studies on chronic pain appear to be limited. This study aims to explore nurses' attitudes toward patients experiencing chronic pain and to provide a contribution to the existing body of literature on the subject.

In this research, the mean total score on the Nurses' attitudes toward patients with Chronic Pain Scale was 3.40±0.46, indicating a moderate level of attitude. This result, while being relatively more positive than the negative attitude findings in Al Omari et al. [31] and the poor attitude reports of Jamal et al. [32], reveals that nurses' attitudes toward pain management are weak, similar to other studies in the literature.[21,33-37] The findings of our study show that nurses' attitudes toward chronic pain management are neither completely negative nor adequate, indicating that nurses' need for education and support in this area continues.

In this research, it was determined that sociodemographic variables such as nurses' age, gender, marital status, and educational status were not effective on attitudes toward chronic pain. This finding shows great consistency with studies in the literature, and it is stated in studies that demographic variables are not effective on nurses' knowledge and attitudes regarding pain management.[9,32,38] However, a significant difference was found between the nurses' sensitivity orientation factor of the attitude scale and gender, and it was determined that male nurses' sensitivity orientation scores were lower than female nurses. This finding is consistent with the systematic review by Ortiz et al. [39] stated that female nurses display more empathetic and patient-centered attitudes, while male nurses show a more distant tendency. This result suggests that misconceptions such as not trusting the patient's pain report and stigmatizing the patient are less common among female healthcare workers and that their attitudes toward patients with chronic pain are strongly positive. In addition, contrary to the finding that general demographic variables are not effective, it shows that the gender factor can create differentiation in specific subdimensions.

In this research, nurses who had received training in chronic pain management were found to have higher scores in overall attitudes toward patients with chronic pain, as well as in the misconception orientation and sensitivity orientation subscales. However, self-perceived competence in chronic pain management did not significantly influence their attitudes, and nearly half of the participants (44.9%) did not consider themselves competent in this area. The literature suggests that healthcare professionals who receive training and feel competent in chronic pain management tend to exhibit more sensitive attitudes toward patients suffering from chronic pain. [27,28,40] When these findings are considered collectively, it appears that receiving education plays a determining role in shaping attitudes toward chronic pain, whereas individual perceptions of self-efficacy may not consistently align with those attitudes.

In this research, it was determined that nurses with 11 years or more of professional experience scored higher on the misconception orientation factor. This finding shows that as nurses gain experience, they trust patients' pain reports more and move away from stigmatizing or false beliefs. As working years increase, it can be thought that nurses develop more conscious and sensitive attitudes toward chronic pain. However, there are also opposite results in the literature. For example, AbuBaker et al. [28] stated that nurses with more than 10 years of experience had weaker pain perception compared to those with 5–10 years of experience. These contradictory findings show that the effect of professional experience on attitude and perception may vary depending not only on duration but also on factors such as training received, clinical environment, and individual approaches.

In this research, it was determined that the working unit was effective on the sensitivity orientation factor and the total score of attitudes toward chronic pain, with nurses working in internal medicine departments having higher scores. This result may be due to more frequent and longer contact with chronic patients in internal medicine departments. The less development of sensitivity toward chronic pain due to care for acute post-operative pain in surgical departments, and the intensive workload, time pressure, and acute intervention-focused approaches in intensive care and emergency departments may be effective on attitudes toward chronic pain. This situation reveals the necessity of unit-specific training. A study indicated that emergency department nurses' attitudes toward pain management were inadequate and that the highest perceived barriers to pain management were overcrowding in the emergency department and workload.[41] Another study states that surgical nurses need special competency development strategies regarding pain management and require different pain management approaches from other units.[42] A study on intensive care nurses indicated that inadequate pain management knowledge due to various barriers and difficulties was insufficient, and unacceptable attitudes emerged.[43] It is stated that negative attitudes toward chronic pain prevent the implementation of pain management guidelines.[44]

In our study, it was determined that the presence of chronic pain in nurses' family members did not affect their attitudes toward patients with chronic pain and subdimension scores. This finding contradicts some studies in the literature. For example, in the study conducted by Martorella et al., [45] it was stated that an individual's or their close environment's chronic pain experience was associated with more positive beliefs and attitudes toward people with chronic pain. This situation suggests that personal experiences may strengthen healthcare professionals' patient-centered approaches by increasing their level of empathy.

Limitations

The findings are limited in their generalizability because the study was conducted only with nurses working at a specific hospital. The fact that information was assessed through a questionnaire poses a limitation in that it is impossible to verify whether participants obtained information from external sources before responding.

CONCLUSION

This research revealed that nurses' attitudes toward patients with chronic pain are at a moderate level, with training and working unit significantly affecting attitude scores. Pain management is directly dependent on nurses' approach in care processes, making their leadership role crucial for improving care quality. Therefore, revising nursing curricula with emphasis on pain education, implementing continuous professional development programs, and integrating evidence-based protocols into clinical practice are recommended as effective strategies to enhance nurses' knowledge and attitudes in pain management. Future studies should explore attitude changes over time, examine differences across healthcare settings, and investigate how attitudes influence actual pain management practices.

DECLARATIONS

Ethics Committee Approval: The study was approved by Erzincan Binali Yıldırım University Ethics Committee (No: 293321-08/04, Date: 06/09/2023).

Informed Consent: Informed consent, both verbal and written, was obtained from the participants who voluntarily agreed to take part in the study.

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: Artificial intelligence (AI)-powered technologies (such as large language models [LLMs], chatbots, or image-generation tools) were not used in the preparation, writing, analysis, or editing of this study. All content in this study was originally prepared by the authors.

Authorship Contributions: Concept – OD, EDS; Design – OD; Supervision – OD, EDS; Fundings – OD, EDS; Materials – OD, EDS; Data collection &/or processing – OD, EDS; Analysis and/or interpretation – OD, EDS; Literature search – OD, EDS; Writing – OD, EDS; Critical review – OD, EDS.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Raja SN, Carr DB, Cohen M, Finnerup NB, Flor H, Gibson S, et al. The revised International Association for the Study of Pain definition of pain: concepts, challenges, and compromises. Pain 2020;161:1976–82.
2. Treede RD, Rief W, Barke A, Aziz Q, Bennett MI, Benoliel R, et al. Chronic pain as a symptom or a disease: the IASP Classification of Chronic Pain for the International Classification of Diseases (ICD-11). Pain 2019;160:19–27.
3. Cohen SP, Vase L, Hooten WM. Chronic pain: an update on burden, best practices, and new advances. Lancet 2021;397:2082–97.
4. The Lancet. Rethinking chronic pain. Lancet 2021;397:2023.
5. Kress HG, Aldington D, Alon E, Coaccioli S, Collett B, Coluzzi F, et al. A holistic approach to chronic pain management that involves all stakeholders: change is needed. Curr Med Res Opin 2015;31:1743–54.
6. International Association for the Study of Pain (IASP). IASP curriculum outline on pain for nursing. 2021. Available from: https://www.iasp-pain.org/education/curricula/iasp-curriculum-outline-on-pain-for-nursing/?ItemNumber=2052. Accepted Aug 26, 2025.
7. Melia R, Morrell-Scott N, Maine N. A review of compliance with pain assessments within a UK ICU. Br J Nurs 2019;28:382–6.
8. Glarcher M, Kundt FS, Meissner W, Osterbrink J. Quality indicators (QI) of acute pain after surgery in european countries. Pain Manag Nurs 2021;22:730–9.
9. Al-Atiyyat N, Salim NA, Tuffaha MG, Abu Nigim HA, Saleh MM, Alkhodary ME, et al. A survey of the knowledge and attitudes of oncology nurses toward pain in United Arab Emirates oncology settings. Pain Manag Nurs 2019;20:276–83.
10. Khalil H, Mashaqbeh M. Areas of knowledge deficit and misconceptions regarding pain among jordanian nurses. Pain Manag Nurs 2019;20:649–55.
11. Damico V, Murano L, Macchi G, Forastieri Molinari A, Dal Molin A. Knowledge and attitude of pain management among Italian nurses in intensive care unit: a multicentric descriptive study. Ann Ig 2021;33:220–30.
12. Moceri JT, Drevdahl DJ. Nurses' knowledge and attitudes toward pain in the emergency department. J Emerg Nurs 2014;40:6–12.
13. Vaismoradi M, Skär L, Söderberg S, Bondas TE. Normalizing suffering: A meta-synthesis of experiences of and perspectives on pain and pain management in nursing homes. Int J Qual Stud Health Well-being 2016;11:31203.
14. Brennan F, Carr D, Cousins M. Access to pain management-still very much a human right. Pain Med 2016;17:1785–9.
15. Pierik JGJ, IJzerman MJ, Gaakeer MI, Vollenbroek-Hutten MMR, Doggen CJM. Painful discrimination in the

emergency department: risk factors for underassessment of patients' pain by nurses. J Emerg Nurs 2017;43:228–38

16. Carvalho AS, Martins Pereira S, Jácomo A, Magalhães S, Araújo J, Hernández-Marrero P, et al. Ethical decision making in pain management: a conceptual framework. J Pain Res 2018;11:967–76.
17. Jukić M, Puljak L. Legal and ethical aspects of pain management. Acta Med Acad 2018;47:18–26.
18. Seers T, Derry S, Seers K, Moore RA. Professionals underestimate patients' pain: a comprehensive review. Pain 2018;159:811–8.
19. Angeletti C, Guetti C, Paesani M, Colavincenzo S, Ciccozzi A, Angeletti PM. An analysis of Italian nurses' approach to patients' pain: a nationwide online survey. Pain Res Manag 2018;2018:5165262.
20. Latina R, Forte P, Mastroianni C, Paterniani A, Mauro L, Fabriani L, et al. Pain education in schools of nursing: a survey of the Italian academic situation. Prof Inferm 2018;70:115–22.
21. Samarkandi OA. Knowledge and attitudes of nurses toward pain management. Saudi J Anaesth 2018;12:220–6.
22. Giusti GD, Reitano B, Gili A. Pain assessment in the Emergency Department. Correlation between pain rated by the patient and by the nurse. An observational study. Acta Biomed 2018;89:64–70.
23. Karra R, Holten-Rossing S, Mohammed D, Parmeggiani L, Heine M, Namnún OC. Unmet needs in the management of functional impairment in patients with chronic pain: a multinational survey. Pain Manag 2021;11:303–14.
24. McKinnon C, White J, Morgan P, Harvey A, Clancy C, Fahey M, et al. Clinician perspectives of chronic pain management in children and adolescents with cerebral palsy and dyskinesia. Phys Occup Ther Pediatr 2021;41:244–58.
25. International Association for the Study of Pain (IASP). Global year for translating pain knowledge to practice. 2022. Available from: https://www.iasp-pain.org/advocacy/global-year/translating-pain-knowledge-to-practice/. Accessed Aug 26, 2025.
26. Oner B, Zengul FD, Oner N, Ivankova NV, Karadag A, Patrician PA. Nursing-sensitive indicators for nursing care: a systematic review (1997-2017). Nurs Open 2021;8:1005–22.
27. Sucu Çakmak NC, Çalışkan N, Kan A. Health care professionals' attitudes toward patients with chronic pain: scale development study. Pain Manag Nurs 2023;24:188–95.
28. AbuBaker NA, Salim NA, Joshua R, Jose A. Registered nurses' perception about pain management of hospitalized patients in one of the governmental hospitals in Dubai, UAE: A cross-sectional study. Dubai Med J 2019;2:1–5.
29. Beck SL, Dunton N, Berry PH, Brant JM, Guo JW, Potter C, et al. Dissemination and implementation of patient-centered indicators of pain care quality and outcomes. Med Care 2019;57:159–66.
30. De Ruddere L, Craig KD. Understanding stigma and chronic pain: a-state-of-the-art review. Pain 2016;157:1607–10.
31. Al Omari D, Alhabahbeh A, Subih M, Aljabery A. Pain management in the older adult: The relationship between nurses' knowledge, attitudes and nurses' practice in Ireland and Jordan. Appl Nurs Res 2021;57:151388.
32. Jamal K, Alameri RA, Alqahtani FM, AlGarni RS, Alamri NA, Elshnawie HA, et al. Knowledge and attitudes of critical care nurses regarding pain management in Saudi Arabia. Med Arch 2023;77:49–55.
33. Shuaib N. Nurses knowledge regarding pain management among patients in critical care units. JCAO 2018;2:106.
34. Mondol S, Muhammad F, Chowdhury A. Nurses'knowledge and practices related to pain assessment in critically ill patients in a selected private hospital in Bangladesh. Int J Community Med Public Health 2018;5:4219–24.
35. Teixeira J, Durão M. Pain assessment in critically ill patients: an integrative literature review. J Nurs Ref 2016:135–41.
36. Fallatah SMA. Pain knowledge and attitude survey among health-care professionals at a university hospital in Saudi Arabia. Saudi J Med Med Sci 2017;5:155–9.
37. Alsaqri SH. Nursing student's knowledge and attitudes toward pain management at Hail University, Saudi Arabia. Int J Adv Appl Sci 2018;5:75–81.
38. Latina R, Mauro L, Mitello L, D'Angelo D, Caputo L, De Marinis MG, et al. Attitude and knowledge of pain management among italian nurses in hospital settings. Pain Manag Nurs 2015;16:959–67.
39. Ortiz MI, Cuevas-Suárez CE, Cariño-Cortés R, Navarrete-Hernández JJ, González-Montiel CA. Nurses knowledge and attitude regarding pain: A systematic review and meta-analysis. Nurse Educ Pract 2022;63:103390.
40. Darawad M, Alnajar MK, Abdalrahim MS, El-Aqoul AM. Cancer pain management at oncology units: comparing knowledge, attitudes and perceived barriers between physicians and nurses. J Cancer Educ 2019;34:366–74.
41. Kahsay DT, Pitkäjärvi M. Emergency nurses´ knowledge, attitude and perceived barriers regarding pain Management in Resource-Limited Settings: cross-sectional study. BMC Nurs 2019;18:56.
42. Atthayasai J, Chatchumni M, Eriksson H, Mazaheri M. Surgical nurses' perceptions of strategies to enhance pain

management proficiency: a qualitative study. Nurs Rep 2023;13:923–33.

43. Almutairi AM, Pandaan IN, Alsufyani AM, Almutiri DR, Alhindi AA, Alhusseinan KS. Managing patients' pain in the intensive care units: Nurses' awareness of pain management. Saudi Med J 2022;43:514–21.
44. Dueñas M, Salazar A, Sánchez M, De Sola H, Ojeda B, Failde I. Relationship between using clinical practice guidelines for pain treatment and physicians' training and attitudes toward patients and the effects on patient care. Pain Pract 2018;18:38–47.
45. Martorella G, Kostic M, Lacasse A, Schluck G, Abbott L. Knowledge, beliefs, and attitudes of emergency nurses toward people with chronic pain. SAGE Open Nurs 2019;5:2377960819871805.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2025.83093
Eur Arch Med Res 2026;42(2):183–187

The Importance of Screen Exposure in Cases Diagnosed with Enuresis

Ozgur Ozdemir Simsek,[1] Gamze Yakin,[2] Hasan Turan,[3] Hasan Akin Tahillioglu[4]

[1]Department of Pediatric Nephrology, Izmir Bakircay University, Faculty of Medicine, Izmir, Turkiye
[2]Department of Pediatrics, Izmir Bakircay University, Faculty of Medicine, Izmir, Turkiye
[3]Department of Pediatric Urology, Izmir Bakircay University, Cigli Training and Research Hospital, Izmir, Turkiye
[4]Department of Child and Adolescent Psychiatry, Izmir Bakircay University, Faculty of Medicine, Izmir, Turkiye

ABSTRACT

Objective: "Screen time" refers to the time spent using all screen-enabled devices throughout the day. When considering sleep patterns and quality, daily screen use is thought to be effective in enuresis due to its impact on cognitive and physical functions. The study aimed to evaluate the relationship between enuresis and screen exposure.

Materials and Methods: One hundred forty-five children aged 6–18 years who applied to Pediatric Nephrology and Pediatric Urology outpatient clinics due to primary enuresis and whose daily screen time was over 60 min were evaluated with questions.

Results: A total of 145 patients, 72 girls and 73 boys, were evaluated with a mean age of 9.3±3.2 years. One hundred twelve patients had a family history of enuresis, and 109 first-degree relatives had a history of enuresis. The children's mean screen exposure time was 5.7±1.6 h/day, and 66% (98 patients) reported screen exposure immediately before sleep. All were observed to have computer or tablet exposure. At 3 months of treatment, all 47 patients with ongoing enuresis were observed to have screen exposure before bedtime, all were observed to have used computers or tablets, and all those with ongoing complaints were observed to have played games. At 6 months of treatment, all 27 patients whose enuresis persisted despite behavioral therapy and medical treatment were observed to have continued screen exposure before bedtime, exceeding 120 min by playing games.

Conclusion: Enuresis is multifactorial, and the effects of screen time are known. In children, in addition to behavioral therapy and medical treatment, reducing screen time, especially gaming time, should be considered in the treatment of enuresis.

Keywords: Enuresis, Gaming time, Screen exposure

Cite this article as: Ozdemir Simsek O, Yakin G, Turan H, Tahillioglu HA.The Importance of Screen Exposure in Cases Diagnosed with Enuresis. Eur Arch Med Res 2026;42(2):183–187.

Address for correspondence: Ozgur Ozdemir Simsek. Department of Pediatric Nephrology, Izmir Bakircay University, Faculty of Medicine, Izmir, Turkiye

E-mail: ozgur_ozdemir_07@hotmail.com **ORCID ID:** 0000-0001-5705-7452

Submitted: 22.09.2025 **Revised:** 23.10.2025 **Accepted:** 11.11.2025 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

INTRODUCTION

The International Pediatric Urinary Incontinence Society (ICCS) defines enuresis as the incontinence of urine during sleep in children aged five and older.[1] Enuresis is classified as monosymptomatic in children without lower urinary tract symptoms (e.g., frequency, daytime incontinence, urgency, genital or lower urinary tract pain) and a history of bladder dysfunction; it is classified as secondary in children with symptoms. Children with monosymptomatic enuresis who have not yet achieved satisfactory dryness are classified as primary; children who develop symptoms after six months or more of dryness are classified as secondary.[1]

Primary monosymptomatic enuresis has a high rate of spontaneous resolution, with prevalence rates decreasing from approximately 15% at age 5–5% at age 10 and from 1% to 2% in individuals aged 15 years and older.[2] However, the longer enuresis lasts, the lower the rate of spontaneous recovery.[3] Enuresis is a clinical condition characterized by the inability to arouse from sleep. It may be caused by one or more factors, including delayed maturation in the child, genetic factors, nocturnal polyuria, sleep disturbance, low bladder capacity, and detrusor overactivity.[4] Psychological and behavioral abnormalities are encountered as both the cause and the consequence of enuresis.[4] Sleep quality, ability to wake up from sleep, and distinctions between REM and non-REM sleep stages are factors that affect enuresis in terms of the sleep-wake cycle.[5] Studies have been conducted to examine the effects of childhood screen exposure on sleep disturbances. The literature has shown that screen use leads to an increase in sleep disturbances and behavioral problems.[6] With the developing world and technology, internet and tablet usage are increasing; the age of usage is decreasing all over the world, and the time spent in front of the screen is increasing day by day.[7] As internet use and time spent online increase, behavioral problems also increase. As patients' sleep patterns and sleep-wake cycles change, increasing eating disorders and inactivity also become a problem.[8] Decreased physical activity and increased obesity are negative parameters in enuresis, leading to increased bladder and intra-abdominal pressure, fluid intake at inappropriate times, and disruption of the antidiuretic hormone secretion cycle.[8,9] Enuresis is a common and treatable condition, and in addition to medical treatment, it is also essential to manage the patient's behavior. In these cases, behavioral changes include regulating hourly water intake, avoiding junk food and eating at set times, and preventing constipation.[8] In addition, research is ongoing to evaluate reducing television, tablet, and internet use before bed as a treatment approach. It is known that all types of screen use affect all types of enuresis diseases and enuresis at all ages.[10,11]

This study aimed to evaluate the use of tablets and television in patients receiving medical treatment for enuresis, to determine the duration of gaming and video viewing, and to assess their impact on treatment outcomes. The aim was to provide evidence-based recommendations to guide patients in monitoring without medication or with shorter-term medication use by recommending behavioral changes before or in conjunction with medical treatment.

MATERIALS AND METHODS

Study Design and Patient Selection

Only cases diagnosed with monosymptomatic enuresis were evaluated. Patients aged 60 months to 18 years who were still receiving treatment but had completed at least 3 months of enuresis treatment between February 2025 and May 2025 were evaluated. Patients who were evaluated in pediatric nephrology and pediatric urology clinics, who were using regular medication, had no known nephrological or urological problems, were not diagnosed with neurodevelopmental or neuropsychiatric diseases, and were not using neurological or psychiatric medications, were evaluated for screen exposure using survey questions. No artificial intelligence (AI)-supported technologies (such as large language models [LLM], chatbots, or image generators) were used in the production of the presented work. This retrospective cross-sectional study was approved by the institutional ethics committee (approval number: 2014, date: February 6, 2025) and was in accordance with the principles of the Declaration of Helsinki. Written consent was obtained from all participants in accordance with ethical rules.

Evaluation

A questionnaire was used to collect information on patients' age at presentation, family history of enuresis, age at toilet training, average sleep duration, difficulty falling asleep due to screen exposure, daily screen time, type of screen exposure (tablet, phone, television), presence of a social media account, time spent gaming, time spent watching videos/movies, type and duration of the last screen exposure before sleep, and whether the previous stimulus was used to induce sleep. Patients were evaluated at the 3rd month of treatment. Patients who still had symptoms at the 3rd month were reevaluated at the 6th month.

Statistical Analysis

No AI-supported technologies (such as LLM, chatbots, or image generators) were used in the production of the presented work. The Statistical Package for the Social Sciences (SPSS) package program (IBM SPSS Statistics for Windows, version 25.0. Armonk, NY: IBM Corp, 2017) was used for statistical analyses. Variables with normal distribution are

shown as mean values±standard deviation, variables with abnormal distribution are shown as median (range), and the rest are expressed as frequency. The Chi-square test was used to compare categorical variables between groups. The Kolmogorov-Smirnov test was used to evaluate the normal distribution of continuous variables between groups. All parameters were distributed abnormally, so they were evaluated by the Mann-Whitney U test.

RESULTS

Three hundred and forty-two patients receiving medical treatment for enuresis were considered. Patients who spent more than 60 min/day in front of screens were identified. A total of 145 patients, 72 women and 73 men, were evaluated via questionnaire. The mean age was 9.3±3.2 years. Of the 145 patients, 112 (77%) had a family history of enuresis, and 109 had a first-degree relative with enuresis. The mean age at which patients began toilet training was 2.5±0.8 years. Patients with symptoms of frequent urination, urgency, urinary incontinence, urinary retention, or recurrent urinary tract infections were excluded from the evaluation. Mean daily sleep duration was 9.4 (7–10.5) h. The average time children spent in front of the screen was found to be 5.7±1.6 h/day. It was observed that all patients were exposed to computer or tablet use, 42 patients had social media accounts, and all boys played games. It was observed that patients who spent 3.5 h or more in front of the screen per day had difficulty falling asleep. Sixty-six percent (98 patients) of patients reported exposure to screens immediately before sleep, and all of these patients had difficulty falling asleep. In the 3rd month of treatment, all 47 patients with persistent enuresis were observed to be exposed to screens before bed. All were using computers or tablets, and all those with persistent complaints were observed to be playing games. In the 6th month of treatment, all 27 patients with persistent enuresis, despite behavioral therapy and medical treatment, were observed to continue to be exposed to screens by playing games before bed.

DISCUSSION

Enuresis treatment is divided into two groups: Medical and paramedical. It is known that the patients who benefit most are those who share a common denominator between these two groups: Those who are treated with both medication and behavioral modification.[12] Even if full compliance with medical treatment is achieved, if behavioral changes are not made, the treatment response will not be sufficient, and the rate of recurrence of the disease will increase at the end of treatment.[13] Enuresis behavioral therapy involves establishing nutritional habits, ensuring adequate fluid intake during daylight hours, meeting physiological sleep needs, maintaining regular sleep patterns, and adhering to appropriate bedtime routines.[14] Prolonged or increased screen use causes children and adolescents to drift away from their daily routine and prevents behavioral changes in treatment.[14] A study examining screen exposure, sleep quality, and behavioral problems in 374 children found that children with screens in their bedrooms experienced more sleep and behavioral issues. The same study also showed that increased exposure, particularly in the evening, led to more sleep disruptions, which were in turn associated with behavioral problems.[6] Screen presence in the bedroom is also associated with obesity in children. Lifestyle behaviors (television viewing time, diet, physical activity, and sleep duration) determine the association with obesity. This is considered a mediating behavior. Reducing screen use has been shown to reduce obesity and behavioral problems.[8] Eating disorders and excessive social media engagement are significant problems among young people. This study assessed the prevalence of eating disorders and their relationship to social media addiction among 350 young people aged 14–25.[15] The American Psychiatric Association recently included Internet Gaming Disorder (IGD) among its potential diagnoses. While the etiology of IGD is not yet fully understood, impulsivity and excessive gaming appear to be risk factors.[7] Another study evaluating 1.713 children showed that screen exposure and sleep duration were inversely proportional, and that increased exposure led to behavioral and adaptation disorders.[12] Experimental studies measuring physiological and polysomnographic information related to screen exposure and sleep disruption have shown that screen light significantly disrupts sleep continuity and quality. In particular, short-wavelength light emitted from monitors has been shown to inhibit or delay normal melatonin production. Light exposure in the evening has been shown to directly affect melatonin's circadian cycle, leading to circadian rhythm disturbances and insomnia, with this effect being more pronounced in adolescents.[16] Children whose parents watched television or played video games in their rooms reported more sleep disturbances. Furthermore, evening screen exposure has been linked to impaired daytime attention and behavior and has been shown to exacerbate sleep disturbances in children ages 6–16. In this digital age, children are often exposed to both screens and activities like internet use and gaming from early childhood. Research shows that young children are particularly vulnerable to screen exposure.[17]

In our study, we found that patients with prolonged screen exposure either did not respond fully to treatment or subsequently experienced relapses. In other words, there was a statistically negative correlation between treatment benefit and screen exposure duration, consistent with the literature.

Study Limitations

This study has several limitations. First, its retrospective design may have led to missing data or selection bias. Second, the study was conducted as a single-center, cross-sectional study, which may limit the generalizability of the findings to other populations. In this context, a limitation of the study is that the questionnaire used to evaluate the patients is not an accepted, well-established tool with proven reliability and validity, but rather a short questionnaire we created ourselves. Information regarding the activation of yellow light or "night mode" filters on the devices was not collected during the study. We are aware that light color temperature may affect sleep quality, but this was not assessed. This could be considered another limitation of our study.

CONCLUSION

Enuresis is a multifactorial pathology, and screen use is known to be a significant factor. Both direct and indirect effects of screen use are known. Literature has shown that it causes cognitive, physical, and neuropsychiatric decline in patients. In addition to behavioral therapy and medical treatment, reducing screen use in children, especially gaming time, should be considered. Considering all this data, our study is cross-sectional, and studies with larger patient numbers and more data are needed.

DECLARATIONS

Ethics Committee Approval: The study was approved by Izmir Bakircay University Ethics Committee (No: 2014, Date: 06/02/2025).

Informed Consent: Informed consent was obtained from all individual participants included in the study.

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: Not declared.

Authorship Contributions: Concept – OOS; Design – OOS; Supervision – OOS; Data collection &/or processing – GY, HT; Analysis and/or interpretation – GY, HT; Literature search – OOS, HAT; Writing – OOS, HAT; Critical review – OOS.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Austin PF, Bauer SB, Bower W, Chase J, Franco I, Hoebeke P, et al. The standardization of terminology of lower urinary tract function in children and adolescents: Update report from the standardization committee of the International Children's Continence Society. Neurourol Urodyn 2016;35:471–81.
2. Forsythe WI, Redmond A. Enuresis and spontaneous cure rate. Study of 1129 enuretis. Arch Dis Child 1974;49:259–63.
3. Bakker E, van Sprundel M, van der Auwera JC, van Gool JD, Wyndaele JJ. Voiding habits and wetting in a population of 4,332 Belgian schoolchildren aged between 10 and 14 years. Scand J Urol Nephrol 2002;36:354–62.
4. Jørgensen CS, Horsdal HT, Rajagopal VM, Grove J, Als TD, Kamperis K, et al. Identification of genetic loci associated with nocturnal enuresis: a genome-wide association study. Lancet Child Adolesc Health 2021;5:201–9.
5. Fernandes AER, Roveda JRC, Fernandes CR, Silva DF, de Oliveira Guimarães IC, Lima EM, de et al. Relationship between nocturnal enuresis and sleep in children and adolescents. Pediatr Nephrol 2023;38:1427–38.
6. Cavalli E, Anders R, Chaussoy L, Herbillon V, Franco P, Putois B. Screen exposure exacerbates ADHD symptoms indirectly through increased sleep disturbance. Sleep Med 2021;83:241–7.
7. Gentile DA, Bailey K, Bavelier D, Brockmyer JF, Cash H, Coyne SM, et al. Internet gaming disorder in children and adolescents. Pediatrics 2017;140(Suppl 2):S81–5.
8. Borghese MM, Tremblay MS, Katzmarzyk PT, Tudor-Locke C, Schuna JM Jr, Leduc G, et al. Mediating role of television time, diet patterns, physical activity and sleep duration in the association between television in the bedroom and adiposity in 10 year-old children. Int J Behav Nutr Phys Act 2015;12:60.
9. Dennison BA, Erb TA, Jenkins PL. Television viewing and television in bedroom associated with overweight risk among low-income preschool children. Pediatrics 2002;109:1028–35.
10. Thompson DA, Christakis DA. The association between television viewing and irregular sleep schedules among children less than 3 years of age. Pediatrics 2005;116:851–6.
11. Marinelli M, Sunyer J, Alvarez-Pedrerol M, Iñiguez C, Torrent M, Vioque J, et al. Hours of television viewing and sleep duration in children: a multicenter birth cohort study. JAMA Pediatr 2014;168:458–64.
12. Neveus T, Eggert P, Evans J, Macedo A, Rittig S, Tekgül S, et al. Evaluation of and treatment for monosymptomatic enuresis: a standardization document from the International Children's Continence Society. J Urol 2010;183:441–7.
13. Oğraş MS, Ağlamış E, Yücel MÖ, Taşdemir C, Selçuk EB. Treatment and follow up results of patients with monosymptomatic enuresis nocturna. Dicle Med J 2013;40:410–3.

14. Madigan S, Browne D, Racine N, Mori C, Tough S. Association between screen time and children's performance on a developmental screening test. JAMA Pediatr 2019;173:244–50. Erratum in: JAMA Pediatr 2019;173:501–2.
15. Mushtaq T, Ashraf S, Hameed H, Irfan A, Shahid M, Kanwal R, et al. Prevalence of eating disorders and their association with social media addiction among youths. Nutrients 2023;15:4687.
16. Nagare R, Plitnick B, Figueiro MG. Effect of exposure duration and light spectra on nighttime melatonin suppression in adolescents and adults. Light Res Technol 2019;51:530–43.
17. Owens J, Maxim R, McGuinn M, Nobile C, Msall M, Alario A. Television-viewing habits and sleep disturbance in school children. Pediatrics 1999;104:e27.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2026.03206
Eur Arch Med Res 2026;42(2):188–196

Machine Learning–Based Ultrasound Radiomics for Differentiating Benign and Malignant Skin Lesions: A Comparative Feature-Set and Model Performance Analysis

Hakan Ayyildiz

Department of Radiology, Basaksehir Cam and Sakura City Hospital, Istanbul, Türkiye

ABSTRACT

Objective: This study aimed to assess the ability of radiomics features extracted from grayscale B-mode dermatologic ultrasound images to distinguish benign from malignant skin lesions and to compare the performance of multiple machine learning (ML) classifiers trained on different radiomics feature domains, including original, Laplacian-of-Gaussian (LoG), and wavelet-transformed images.

Materials and Methods: This retrospective study analyzed 190 skin lesions (129 benign, 61 malignant) from a publicly available dermatologic ultrasound dataset. Manual lesion segmentation was performed by an experienced radiologist. Radiomics features were extracted from original grayscale, LoG-filtered, and wavelet-transformed images using PyRadiomics. Low-variance descriptors and highly correlated features were removed, and mutual information–based feature selection was conducted within stratified five-fold cross-validation to avoid information leakage. Logistic regression, support vector machine (RBF), and random forest classifiers were trained using balanced class weights. Performance metrics were computed by pooling predictions from all folds.

Results: A total of 851 radiomics features per lesion were initially extracted and reduced to 143 non-redundant descriptors after preprocessing. Across all models and feature domains, the area under the receiver operating characteristic curve (AUC) ranged from 0.64 to 0.71. The highest performance was achieved by the random forest classifier using the combined feature set (AUC: 0.706; accuracy: 0.716). Wavelet-only features also performed well, with random forest achieving an AUC of 0.705 and the highest accuracy (0.742). Logistic regression applied to original features yielded an AUC of 0.697 and comparatively better sensitivity. Specificity was consistently higher than sensitivity across classifiers.

Conclusion: Radiomics-based analysis of dermatologic ultrasound provides moderate discriminatory ability for benign–malignant classification. Although limited by dataset size and manual segmentation, the findings support radiomics as a quantitative, operator-independent adjunct for ultrasound-based skin lesion evaluation.

Keywords: Machine Learning, Radiomics, Skin Neoplasms, Ultrasonography

Cite this article as: Ayyildiz H. Machine Learning–Based Ultrasound Radiomics for Differentiating Benign and Malignant Skin Lesions: A Comparative Feature-Set and Model Performance Analysis. Eur Arch Med Res 2026;42(2):188–196.

Address for correspondence: Hakan Ayyildiz. Department of Radiology, Basaksehir Cam and Sakura City Hospital, Istanbul, Türkiye
E-mail: hakanayyildiz77@gmail.com **ORCID ID:** 0000-0003-0244-7765

Submitted: 27.12.2025 **Revised:** 11.01.2026 **Accepted:** 12.01.2026 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

INTRODUCTION

Skin cancer is one of the most rapidly increasing malignancies worldwide, and early diagnosis remains critical for improving patient outcomes.[1] Although dermoscopy and clinical examination are widely used in the initial evaluation of suspicious lesions, imaging modalities such as ultrasound (US) have gained increasing importance due to their accessibility, lack of ionizing radiation, real-time capability, and ability to visualize subcutaneous extension.[2] High-frequency ultrasound, in particular, provides valuable morphological and structural information that assists clinicians in differentiating benign from malignant cutaneous lesions.[3] However, sonographic assessment is inherently operator-dependent, and subtle textural patterns that may be diagnostic are often beyond the threshold of human perception.[3]

Radiomics offers a solution to these limitations by converting medical images into high-dimensional, mineable quantitative data.[4] Through the extraction of texture, intensity, and wavelet-based features, radiomics can reveal microstructural heterogeneity that correlates with the biological characteristics of tumors. Although radiomics has been extensively studied in computed tomography and magnetic resonance imaging, the number of studies applying radiomics analysis to ultrasound for skin lesions remains very limited.[5] The acoustic properties and speckle patterns inherent to ultrasound present a unique opportunity for texture-based characterization but also introduce technical challenges that require careful feature extraction and model optimization.[6]

In recent years, machine learning (ML) algorithms have demonstrated promising performance across various oncologic imaging tasks.[7] However, the optimal feature domains and model types for ultrasound-based skin lesion radiomics remain unclear. Prior studies have typically focused on handcrafted features from original images without systematically exploring the impact of filtered feature sets, such as Laplacian-of-Gaussian (LoG) or wavelet transformations, which may provide additional discriminatory power by emphasizing different textural scales.[8] Moreover, comparative evaluations of multiple machine learning classifiers applied to different radiomics feature domains in skin lesion ultrasound imaging are scarce.[9]

Given these gaps, there is a need for a comprehensive analysis that examines how different radiomics feature sets and machine learning models influence diagnostic performance in distinguishing benign from malignant skin lesions. Such an approach would not only provide insights into the most informative texture domains of ultrasound images but also highlight which ML methods are most robust for this classification task.

Therefore, this study aimed to evaluate and compare the performance of multiple radiomics feature groups, including original, Laplacian-of-Gaussian-filtered, and wavelet-transformed features, combined with different machine learning classifiers for differentiating benign and malignant skin lesions on ultrasound. By systematically assessing these combinations using a standardized cross-validated framework, we sought to identify effective radiomics–machine learning strategies for non-invasive lesion characterization.

MATERIALS AND METHODS

Dataset and Study Design

This retrospective radiomics analysis was performed using an openly available dermatologic ultrasound dataset published by Laverde-Saad et al.,[10] and hosted on the Kaggle platform.[11] The dataset consists of grayscale B-mode and color Doppler dermatologic ultrasound images acquired according to DERMUS guidelines. A total of 235 surgically excised or biopsied skin lesions were scanned, and seven diagnostic categories representing the most common benign and malignant cutaneous lesions were included. All images were de-identified before release and contained no patient-identifying information. For the present study, grayscale B-mode images and corresponding lesion masks were used; cases without grayscale B-mode ultrasound images or complete data were excluded, resulting in 190 lesions (129 benign, 61 malignant) for final analysis.

All ultrasound scans in the source dataset were acquired using a single ultrasound platform (MyLab Class C, Esaote, Geneva) with 10–22 MHz or 18 MHz linear probes to maintain image homogeneity, as described in the original publication.[10]

The images were obtained from a publicly available and fully de-identified dataset; therefore, no institutional ethical approval or informed consent was required for this study.

Lesion Segmentation

All images were segmented manually by a radiologist with 8 years of experience in diagnostic ultrasound. Segmentation was performed in a blinded manner, without any knowledge of the histopathological diagnosis (benign or malignant) of the lesions, to prevent reader bias.

Segmentation was carried out using 3D Slicer software. A freehand region of interest (ROI) was drawn on the grayscale B-mode ultrasound image containing the largest cross-sectional diameter of the lesion. The ROI was carefully adjusted to include the entire lesion while excluding adjacent normal tissues or artifacts. Segmentation masks were exported in NRRD labelmap format and reviewed for accuracy before radiomics feature extraction. The radiomics feature domains used in this study are illustrated in Figure 1, which

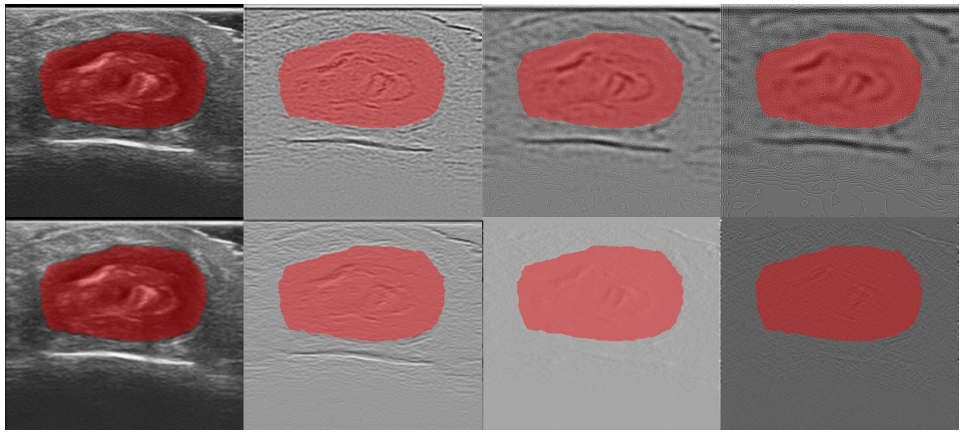

Figure 1. Visualization of multi-domain radiomics filters applied to a grayscale ultrasound image of a cutaneous lesion. The first row demonstrates the original image and Laplacian of Gaussian (LoG) filters with σ = 2, 4, and 6. The second row shows single-level Haar wavelet decompositions (LL, LH, HL, HH). The manually-drawn segmentation mask is overlaid on all images to illustrate the correspondence of texture-derived features within the lesion boundaries.

displays the original ultrasound image, LoG filters (σ=2, 4, 6), and Haar wavelet sub-bands (LL, LH, HL, HH) with the lesion segmentation overlaid.

Radiomics Feature Extraction

Radiomics features were extracted using the PyRadiomics framework implemented in Python. Before feature extraction, images were preprocessed using the default PyRadiomics pipeline: Z-score normalization within the ROI, gray-level discretization using a bin width of 10, B-spline interpolation, and binary mask thresholding to maintain segmentation integrity.

To investigate the impact of different texture domains, radiomics features were calculated from original grayscale images, Laplacian-of-Gaussian (LoG)-filtered images, and wavelet-transformed images (eight directional decompositions).

A total of 851 features per lesion were initially extracted, including first-order statistics, shape features, gray-level co-occurrence matrix (GLCM), gray-level run-length matrix (GLRLM), gray-level size-zone matrix (GLSZM), gray-level dependence matrix (GLDM), and neighboring gray-tone difference matrix (NGTDM) features. Automatic diagnostic metadata generated by PyRadiomics were excluded.

Feature Processing and Machine Learning Workflow

Following feature extraction, all radiomics variables underwent a structured preprocessing workflow to ensure robustness and reduce dimensionality. Initially, the dataset was examined for non-informative descriptors, and features with zero variance were removed. This step eliminated attributes that did not contribute measurable variability between lesions. Subsequently, redundancy among the remaining features was addressed by constructing a Pearson correlation matrix. When two features demonstrated a correlation coefficient exceeding 0.85, one of them was discarded to prevent multicollinearity and ensure a more stable learning environment. After this unsupervised reduction process, three distinct feature groups were retained for modeling: original grayscale features, wavelet-transformed features, and a combined feature set incorporating both domains.

Statistical Analysis and Model Evaluation

Machine learning analyses were then conducted using Python-based pipelines. To avoid information leakage, all supervised feature selection procedures were embedded strictly within the training folds of the cross-validation framework. Mutual information (SelectKBest) was employed to identify the most informative features, with a maximum of 30 variables selected in each fold, depending on feature availability after unsupervised filtering. This ensured that each classifier was trained on an optimized subset of radiomics descriptors without exposing the test folds to any stage of the feature selection process.

Model construction was performed using three commonly applied algorithms in radiomics research: logistic regression with L2 regularization, support vector machine with a radial

basis function kernel, and random forest classifier with 500 trees. Because of the class imbalance between benign and malignant lesions, all models were trained using balanced class weights to prevent bias toward the majority class. Model performance was evaluated using five-fold stratified cross-validation, allowing malignant lesions to be proportionally represented across folds. Probabilistic outputs from all folds were aggregated to compute pooled performance metrics, including accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve. By assessing each combination of feature group and classifier, the workflow enabled a comprehensive comparison to identify the most discriminative radiomics-based approach for distinguishing benign from malignant skin lesions. A visual summary of the overall workflow, including dataset curation, segmentation, feature extraction, and model evaluation, is presented in Figure 2.

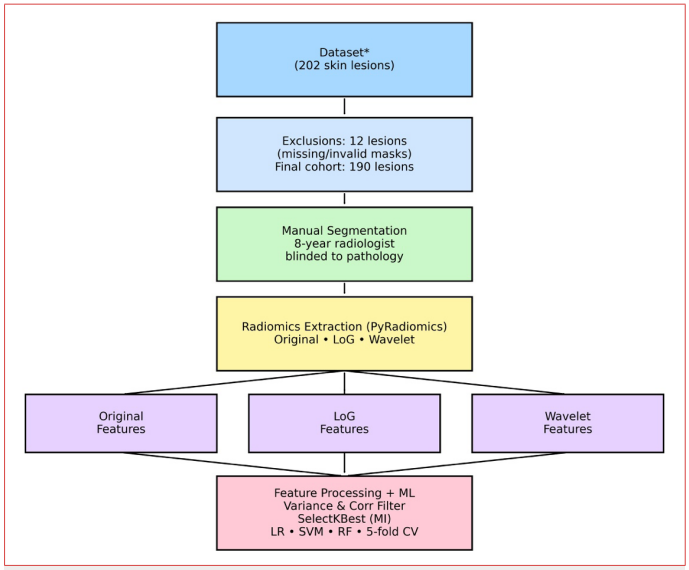

Figure 2. Workflow of the radiomics pipeline, including dataset selection, segmentation, radiomics feature extraction (original, LoG, and wavelet), and machine learning model development.

RESULTS

Dataset Characteristics

A total of 202 dermatologic ultrasound lesions were initially reviewed. Twelve lesions were excluded due to missing or invalid segmentation masks, yielding a final cohort of 190 analyzable lesions, including 129 benign and 61 malignant cases. All remaining lesions successfully underwent radiomics processing.

Radiomics Feature Extraction and Reduction

Radiomics extraction from the original, LoG, and wavelet-transformed images produced 851 features per lesion. The dimensionality of this feature set was subsequently reduced through a multistep process. Low-variance filtering removed two near-constant features, followed by Pearson correlation filtering, which eliminated highly collinear variables and reduced the dataset to 143 non-redundant features. Further refinement was performed inside the training folds of a stratified five-fold cross-validation framework using mutual information–based SelectKBest, resulting in the selection of the top 30 discriminative features for each classifier during model training.

Machine Learning Model Performance

Three classifiers—logistic regression, a support vector machine with an RBF kernel, and random forest—were trained on the original, LoG, wavelet, and combined feature sets using stratified five-fold cross-validation. Model performance varied across feature domains and algorithms. The highest overall performance was observed with the random forest model trained on the combined feature set, which achieved an AUC of 0.706 and an accuracy of 0.716. This model demonstrated high specificity (0.891) for benign lesions but more modest sensitivity (0.344) for malignant lesions.

Wavelet-derived features also showed strong discriminative value. The random forest model trained solely on wavelet features reached an AUC of 0.705 and the highest accuracy among all models at 0.742. Original features yielded intermediate performance, with logistic regression achieving an AUC of approximately 0.697 and offering relatively better malignant sensitivity compared with the other models. SVM-based models consistently showed high specificity but lower sensitivity, indicating stronger performance in identifying benign lesions. Table 1 summarizes model performance across classifiers and feature sets, while Figure 3 presents the corresponding ROC curves for each feature domain and classifier.

Overall Classification Trends

Across all model–feature combinations, AUC values ranged from 0.64 to 0.71, and accuracy ranged from 0.63 to 0.74. A consistent pattern was observed in which specificity exceeded sensitivity across classifiers, reflecting better performance in correctly identifying benign lesions than malignant ones. Despite this imbalance, the stable results across feature groups and algorithms support the conclusion that quantitative radiomics features derived from dermatologic ultrasound images contain diagnostically relevant information for benign/malignant differentiation (Table 1).

DISCUSSION

In this study, we demonstrated the feasibility of using radiomics analysis of grayscale B-mode ultrasound images to

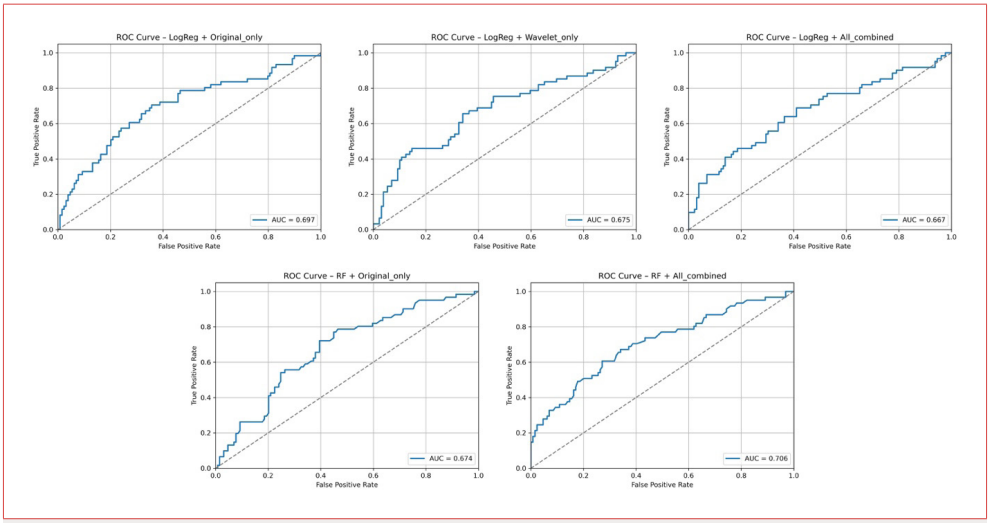

Figure 3. Receiver operating characteristic (ROC) curves for logistic regression and random forest models across different radiomics feature domains. The top row displays ROC performance for logistic regression using original features, wavelet features, and all combined features. The bottom row shows corresponding ROC curves for random forest using original and all combined feature sets. The area under the curve (AUC) is reported for each model.

classify skin lesions as benign or malignant. Using a dataset of approximately 190 ultrasound images, we extracted a high-dimensional set of quantitative imaging features and built machine learning models to distinguish malignant lesions. Our accuracy is comparable to the ~77% test accuracy reported for a deep learning classifier on a similar image set and, notably, is in the same range as expert dermatologist performance.[10] Their deep learning model achieved 77.1% accuracy in differentiating benign from malignant lesions, which was not significantly different from the performance of an experienced dermatologist (74.1%) on the same cases. Thus, our radiomics approach performs on par with advanced deep learning and expert interpretation, underscoring that ultrasound radiomics can capture diagnostically relevant information in skin lesions. This comparison is encouraging because it indicates that the diagnostic signal present in skin ultrasound images can be extracted through different methodological approaches. In fact, our radiomics model slightly exceeded the dermatologist-level accuracy reported by Laverde-Saad et al.,[10] albeit within a similar range. Given that the radiomics and CNN models were tested on comparable images, our study provides evidence that machine learning based on engineered radiomics features is a feasible alternative to deep learning for this task.

It is noteworthy that the random forest classifier performed best when using wavelet-transformed features alone. Wavelet radiomics features likely captured multiscale textural patterns and vascular detail in the ultrasound images, which may be particularly indicative of malignancy. This finding aligns with radiomics studies in other domains showing that multifrequency texture features can enhance lesion discrimination.[12]

Table 1. Performance of Radiomics-Based Machine Learning Models for Classifying Benign and Malignant Skin Lesions on Ultrasound (SVM: Support Vector Machine RBF:Radial Basis Function)

Radiomics Feature Set	Classifier	AUC	Accuracy	Sensitivity	Specificity
All features	Random Forest	0.706	71.6%	34.4%	89.1%
Wavelet features	Random Forest	0.705	74.2%	40.9%	89.9%
Original features	Logistic Regression	0.697	66.8%	62.3%	68.9%
All features	SVM (RBF Kernel)	0.687	73.2%	29.5%	93.8%
Original features	Random Forest	0.674	66.3%	26.2%	85.3%

Our radiomics-based ultrasound classification of skin lesions demonstrates competitive diagnostic performance in line with recent literature. For example, a 2025 study by Qin et al.[8] extracted grayscale ultrasound radiomics features and used LASSO-selected inputs for machine learning, achieving a high AUC of ~0.96 in distinguishing benign from malignant skin tumors. The higher AUC in studies such as that by Qin et al.[8] may reflect differences in dataset homogeneity, lesion types, or optimization strategies, whereas our real-world dataset presents a broader and more challenging diagnostic spectrum. This is comparable to state-of-the-art deep learning models: a 2025 CNN approach combining B-mode and Doppler high-frequency ultrasound images reached approximately 95% accuracy (AUC ~0.98) in binary lesion classification.[13] In a broader prospective study, a deep multimodal network that fused clinical photographs with ultrasound slightly trailed those figures (binary AUC ~0.88), but its performance nearly matched that of expert dermatologists.[14] It should be noted that these performance differences likely reflect variations in model complexity, dataset size, and study design; therefore, direct comparison of AUC values across radiomics- and deep learning–based studies should be interpreted with caution. Accordingly, the contribution of this study lies not in outperforming deep learning models but in providing a systematic and transparent comparison of radiomics feature domains and machine learning classifiers for ultrasound-based lesion characterization.

Overall, both radiomics-based models and modern CNNs have achieved strong accuracy and AUC in dermatologic ultrasound, often markedly outperforming diagnoses based only on clinical features.[8] Key differences lie in feature extraction and modeling. Our radiomics pipeline relies on manual ROI delineation and predefined features, which are statistically filtered and fed into classifiers.[8] This approach can be advantageous for small datasets and offers some interpretability, but it may miss subtle imaging patterns. In contrast, deep learning methods automatically learn abstract features from raw ultrasound images; recent studies have trained CNNs on single-modality ultrasound or dual inputs to capture more information.[13] Others have integrated ultrasound with clinical photographs in a fusion model to improve the classification of skin diseases.[14] Hybrid strategies are also emerging; one 2024 study combined clinical data, radiomics features, and CNN-extracted features using an SVM, yielding an AUC of ~0.95 in external test sets.[15] In comparison, our radiomics and ML method remains robust and relatively data-efficient, but it may be outperformed by deep learning in multiclass or highly varied lesion scenarios. As a trade-off, radiomics models have proven their value by significantly improving diagnostic accuracy over clinical assessment alone, whereas deep learning approaches offer higher automation and the ability to leverage richer image information at the cost of greater complexity.[8] This approach allows direct inspection of feature categories and their relative contributions, offering greater transparency and interpretability compared with deep learning models, which often function as "black boxes." In addition, radiomics-based methods can be more data-efficient and robust in smaller datasets, where training complex deep learning architectures may be impractical or prone to overfitting. Each approach has complementary strengths, and the handcrafted feature model used in this study can be viewed as a solid, interpretable baseline against which newer deep learning or fusion techniques achieve incremental gains in diagnostic performance. Accordingly, the proposed models should be viewed as decision-support tools rather than standalone screening systems, and their outputs must be interpreted in conjunction with clinical and imaging findings.

Several limitations likely contributed to the moderate performance and constrain the generalizability of our findings. First, the dataset was relatively small, which limits lesion diversity and may predispose the models to overfitting or dataset-specific bias. This restricts the generalizability of the results, particularly across different institutions, ultrasound devices, and acquisition protocols. No external validation was performed; therefore, it remains unknown how well the radiomics model would perform on independent data from other institutions or ultrasound devices. This lack of external testing is a critical weakness, as models that perform adequately on internal data can falter when faced with variations in imaging settings or patient populations.

Additionally, our workflow relied on manual lesion segmentation performed by a single radiologist with 8 years of experience, an inherently time-consuming process that can introduce interoperator variability. Interobserver variability was not formally assessed, which represents a limitation of this study. In fact, the requirement for precise ROI segmentation is a known bottleneck in radiomics; it is cumbersome and highly dependent on the operator's experience.[16] This dependency not only affects the reproducibility of the features but also contrasts with deep learning models, which can learn directly from image data without a separate segmentation step.

Taken together, the limited sample size, single-center data, lack of external validation, and reliance on manual segmentation indicate that our reported AUC of ~0.71 should be interpreted with caution. The dataset also showed class imbalance (129 benign vs. 61 malignant lesions), which likely contributed to the relatively low sensitivity observed across models. SMOTE oversampling was evaluated within the training folds and improved sensitivity but resulted in reduced specificity and a slight decrease in AUC; therefore, class-weighted non-SMOTE

models were retained as the primary analysis, with SMOTE results reported in the supplementary material (Suppl 1). The model's generalizability is likely moderate at best, and further work is needed before it can be considered for clinical use.

Despite these limitations, our study highlights that radiomics analysis of high-frequency skin ultrasound is a viable approach for lesion characterization. The performance, although not superior to deep learning benchmarks, confirms that handcrafted ultrasound features contain relevant diagnostic information. With improvements, such as larger multicenter datasets, automated or standardized segmentation, and prospective validation, the accuracy of radiomics-based models may be enhanced. Future investigations might also explore hybrid approaches, such as combining radiomics features with deep learning or clinical data, to improve performance.

CONCLUSION

Radiomics proved moderately effective in classifying dermatologic ultrasound lesions as benign or malignant, providing a foundation for quantitative image analysis in this emerging field. Achieving the higher accuracy levels reported with deep learning (AUC ~0.77) will require further research and validation, given the current model's limitations in data scope and generalizability. Our findings encourage ongoing efforts to refine radiomics and integrate it with advanced AI techniques so that non-invasive ultrasound can be used more reliably as a diagnostic tool in dermatologic oncology.

DECLARATIONS

Ethics Committee Approval: The images were obtained from a publicly available dataset in which all identifying information has been completely anonymized; therefore, no institutional ethics committee approval was required for this study.

Informed Consent: The images were obtained from a publicly available dataset in which all identifying information has been completely anonymized; therefore, no informed consent was required for this study.

Conflict of Interest: The author declare that there is no conflict of interest.

Funding: The author received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: Not declared.

Authorship Contributions: Concept – HA; Design – HA; Supervision – HA; Fundings – HA; Materials – HA; Data collection &/or processing – HA; Analysis and/or interpretation – HA; Literature search – HA; Writing – HA; Critical review – HA.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Rigel DS, Friedman RJ, Kopf AW. The incidence of malignant melanoma in the United States: issues as we approach the 21st century. J Am Acad Dermatol 1996;34:839–47.
2. Raza S, Ali F, Al-Niaimi F. Ultrasonography in diagnostic dermatology: a primer for clinicians. Arch Dermatol Res 2023;315:1–6.
3. Boostani M, Wortsman X, Pellacani G, Cantisani C, Suppa M, Mohos A, et al. Dermoscopy-guided high-frequency ultrasound for preoperative assessment of basal cell carcinoma lateral margins: a pilot study. Br J Dermatol 2025;193:572–4.
4. Rogers W, Thulasi Seetha S, Refaee TAG, Lieverse RIY, Granzier RWY, Ibrahim A, et al. Radiomics: from qualitative to quantitative imaging. Br J Radiol 2020;93:20190948.
5. Falcone R, Verkhovskaia S, Di Pietro FR, Scianni C, Poti G, Morelli MF, et al. Application of radiomics in melanoma: a systematic review and meta-analysis. Cancers (Basel) 2025;17:3130.
6. Thomson H, Yang S, Cochran S. Machine learning-enabled quantitative ultrasound techniques for tissue differentiation. J Med Ultrason (2001) 2022;49:517–28.
7. Zhang YP, Zhang XY, Cheng YT, Li B, Teng XZ, Zhang J, et al. Artificial intelligence-driven radiomics study in cancer: the role of feature engineering and modeling. Mil Med Res 2023;10:22.
8. Qin Y, Zhang Z, Qu X, Liu W, Yan Y, Huang Y. A machine learning model based on high-frequency ultrasound for differentiating benign and malignant skin tumors. Med Ultrason 2025;27:284–93.
9. Faita F, Oranges T, Di Lascio N, Ciompi F, Vitali S, Aringhieri G, et al. Ultra-high-frequency ultrasound and machine learning approaches for the differential diagnosis of melanocytic lesions. Exp Dermatol 2022;31:94–8.
10. Laverde-Saad A, Jfri A, García R, Salgüero I, Martínez C, Cembrero H, et al. Discriminative deep learning based benignity/malignancy diagnosis of dermatologic ultrasound skin lesions with pretrained artificial intelligence architecture. Skin Res Technol 2022;28:35–9.
11. Dermatologic Ultrasound Images for Classification Dataset. Available from: https://www.kaggle.com/datasets/alfageme/dermatologic-ultrasound-images. Accessed Nov 24, 2025.
12. Lee J, Lee J, Song BI. A machine learning-based radiomics model for the differential diagnosis of benign and malignant thyroid nodules in F-18 FDG PET/CT: external validation in the different scanner. Cancers (Basel) 2025;17:331.
13. da Silva IRV, Jardim AG, de Souza Faés GR, Alva TAP, Becker CDL, Botelho VR. Deep learning-based

skin lesion classification: a CNN approach on high-frequency ultrasound imaging. J Ultrasound Med 2026;45:925–36.

14. Zhu AQ, Wang Q, Shi YL, Ren WW, Cao X, Ren TT, et al. A deep learning fusion network trained with clinical and high-frequency ultrasound images in the multi-classification of skin diseases in comparison with dermatologists: a prospective and multicenter study. EClinicalMedicine 2024;67:102391.
15. Dai X, Lu H, Wang X, Zhao B, Liu Z, Sun T, et al. Development of ultrasound-based clinical, radiomics and deep learning fusion models for the diagnosis of benign and malignant soft tissue tumors. Front Oncol 2024;14:1443029.
16. Zhang X, Zhang Y, Zhang G, Qiu X, Tan W, Yin X, Liao L. Deep learning with radiomics for disease diagnosis and treatment: challenges and potential. Front Oncol 2022;12:773840.

Appendix 1. Effect of SMOTE on Random Forest model performance across key radiomics feature sets

Feature Set	Variant	AUC	Accuracy (%)	Sensitivity (%)	Specificity (%)
All combined	No SMOTE	0.706	71.6	34.4	89.1
All combined	SMOTE	0.688	71.1	47.5	82.2
Wavelet only	No SMOTE	0.705	74.2	40.9	89.9
Wavelet only	SMOTE	0.709	71.6	50.8	81.4

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2026.46693
Eur Arch Med Res 2026;42(2):197–205

Impact of Deep Learning–Based Real-Time Polyp Detection on Adenoma Detection Rate and Concurrent Endoscopist-Performed Optical Biopsy Accuracy: Clinical Validation in Routine Colonoscopy

Yavuz Ozden[1], Ferhat Omurca[1], Dilek Tekis[1], Oktay Bulur[2], Nuh Mehmet Buyukberber[1]

[1]Department of Gastroenterology, Kayseri City Hospital, University of Health Sciences, Kayseri, Türkiye
[2]Department of Gastroenterology, Istanbul Medipol University, Istanbul, Türkiye

ABSTRACT

Objective: Adenoma detection rate (ADR) is the strongest quality indicator of colonoscopy and is inversely associated with interval colorectal cancer risk. Although computer-aided detection (CADe) improves ADR in randomized trials, evidence from routine practice and its impact on concurrent endoscopist-performed optical biopsy remains limited. This study evaluated whether real-time deep learning–based CADe improves ADR in routine colonoscopy and assessed the diagnostic performance of endoscopist-performed optical biopsy.

Materials and Methods: This non-randomized prospective pragmatic study was conducted at a tertiary referral center in Türkiye between October 2025 and February 2026. A total of 2,122 consecutive colonoscopies were allocated by weekly alternation to AI-assisted detection (CADe-on, n=1,061) or standard colonoscopy (CADe-off, n=1,061). The CADe system was used exclusively for lesion detection, whereas optical diagnosis was performed by the endoscopist. The primary outcome was ADR; secondary outcomes included polyp detection rate (PDR), adenomas per colonoscopy (APC), lesion characteristics, withdrawal time, and optical biopsy accuracy using histopathology as the reference standard.

Results: Baseline characteristics were comparable between groups. ADR increased from 13.0% to 19.0% (absolute increase, 6% points; p<0.001), and AI assistance remained independently associated with adenoma detection (adjusted odds ratio 1.56; 95% confidence interval [CI] 1.22–1.99). PDR and APC were also higher in the AI-assisted group (35.4% vs. 28.7%, p=0.002; 0.32±0.71 vs. 0.21±0.54, p=0.004). Detection gains were mainly driven by diminutive and non-polypoid lesions, while withdrawal time did not differ significantly (p=0.11). Endoscopist-performed optical biopsy (428 polyps) showed sensitivity 88.4%, specificity 81.2%, positive predictive value 84.7%, negative predictive value (NPV) 85.4%, and overall accuracy 85.1% (95% CI 81.4–88.3). For diminutive rectosigmoid polyps, NPV was 91.2%, meeting the preservation and ıncorporation of valuable endoscopic ınnovations threshold.

Conclusion: Real-time CADe was associated with higher adenoma detection without prolonging withdrawal time. Endoscopist-performed optical characterization during AI-assisted colonoscopy showed reliable diagnostic performance, suggesting that this combined approach may be feasible in routine clinical practice.

Keywords: Adenoma detection rate, Artificial intelligence, Colonoscopy, Computer-aided detection, Deep learning

Cite this article as: Ozden Y, Omurca F, Tekis D, Bulur O, Buyukberber NM. Impact of Deep Learning–Based Real-Time Polyp Detection on Adenoma Detection Rate and Concurrent Endoscopist-Performed Optical Biopsy Accuracy: Clinical Validation in Routine Colonoscopy. Eur Arch Med Res 2026;42(2):197–205.

Address for correspondence: Yavuz Ozden. Department of Gastroenterology, Kayseri City Hospital, University of Health Sciences, Kayseri, Türkiye

E-mail: yavuzozden@gmail.com **ORCID ID:** 0000-0002-5760-3529

Submitted: 06.03.2026 **Revised:** 21.04.2026 **Accepted:** 30.04.2026 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

INTRODUCTION

Colorectal cancer (CRC) remains a major cause of cancer-related morbidity and mortality worldwide despite the widespread implementation of screening programs.[1] Colonoscopy reduces both CRC incidence and mortality primarily through the detection and removal of adenomatous polyps. However, the protective effect of colonoscopy depends critically on the quality of mucosal inspection. The adenoma detection rate (ADR) has emerged as the most widely accepted quality indicator of colonoscopy performance and demonstrates a strong inverse association with the risk of interval CRC.[2] Even modest absolute improvements in ADR translate into clinically meaningful reductions in post-colonoscopy cancer risk.

Despite advances in high-definition imaging and enhanced visualization technologies, adenomas continue to be missed in a considerable proportion of examinations. Tandem colonoscopy studies consistently report miss rates ranging from 20% to 26%, with diminutive and non-polypoid (flat) lesions accounting for a disproportionate share of overlooked neoplasms.[3] These limitations have stimulated the development of artificial intelligence (AI)-based computer-aided detection (CADe) systems designed to assist endoscopists in real time. A growing body of randomized controlled trials and subsequent meta-analyses has shown that CADe increases ADR and reduces adenoma miss rates compared with standard colonoscopy.[4-6] However, most of the available evidence originates from controlled efficacy trials. Whether these benefits persist under the heterogeneous conditions of routine clinical practice – where bowel preparation quality, patient characteristics, and endoscopist experience vary substantially – remains uncertain.

In parallel with advances in lesion detection, real-time optical characterization remains an important component of colonoscopy practice. When sufficiently accurate, optical biopsy may facilitate risk-stratified management strategies such as the "resect-and-discard" or "leave-*in situ*" approach for diminutive rectosigmoid polyps.[7] Nevertheless, the performance of optical diagnosis performed by endoscopists in routine clinical settings, which involves considerable variability in bowel preparation, lesion morphology, and operator experience, remains incompletely characterized, particularly when combined with concurrent AI-assisted detection.

To date, prospective studies evaluating the combined clinical impact of integrating both real-time AI-assisted polyp detection and endoscopist-performed optical biopsy into routine colonoscopy practice are scarce. In particular, it remains unclear whether implementation outside controlled research environments can simultaneously improve adenoma detection while maintaining reliable histologic prediction by the endoscopist.

Therefore, the aim of this prospective pragmatic comparative study was twofold: first, to determine whether real-time deep learning–based CADe improves ADR in routine clinical practice; and second, to evaluate the diagnostic accuracy of endoscopist-performed optical biopsy during AI-assisted colonoscopy using histopathology as the reference standard.

MATERIALS AND METHODS

Study Design and Setting

This non-randomized prospective pragmatic comparative study was conducted at Kayseri City Hospital, a tertiary referral center in Türkiye, between October 01, 2025, and February 01, 2026. The study evaluated the clinical impact of integrating a deep learning–based real-time CADe system into routine colonoscopy practice.

To preserve routine clinical workflow while minimizing allocation bias, allocation to AI-assisted or standard colonoscopy was performed by weekly alternation rather than individual patient randomization. During AI-on weeks, the CADe platform was connected to the endoscopy video output, whereas during AI-off weeks, it remained disconnected. Consecutive patients undergoing colonoscopy within a given week were therefore allocated to the corresponding study arm. This design maintained routine scheduling while facilitating balanced distribution of patient characteristics and endoscopist workload across study periods. The distribution of endoscopists and colonoscopy indications was monitored throughout the study to ensure a comparable case mix between groups. No changes in endoscopic equipment, staffing, or bowel preparation protocols occurred during the study period.

Ethics Approval

The study protocol was approved by the Institutional Ethics Committee (Kayseri City Hospital Non-interventional Clinical Research Ethics Committee; Approval No: 585; September 26, 2025). All procedures were conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants before colonoscopy.

Study Population and Eligibility Criteria

Consecutive adult patients (≥18 years) undergoing elective colonoscopy for screening, surveillance, or diagnostic indications during the study period were eligible for inclusion.

Exclusion criteria were:

- Known inflammatory bowel disease

- Known CRC
- Hereditary CRC syndromes
- Prior colorectal resection
- Emergency colonoscopy.

Bowel preparation quality was assessed using the Boston Bowel Preparation Scale (BBPS) and recorded as a baseline covariate.[8] No patient was excluded solely on the basis of BBPS score, because bowel preparation quality was analyzed as a clinical variable in the primary comparative analysis. Baseline characteristics included age, sex, colonoscopy indication, bowel preparation quality, and withdrawal time.

AI-Assisted Colonoscopy Protocol

All colonoscopies were performed using high-definition colonoscopes (ELUXEO endoscopy system with EC-760Z-V/L colonoscopes, Fujifilm, Tokyo, Japan) in standard endoscopy units by one of five experienced gastroenterologists. Each endoscopist had more than 5 years of independent practice and had performed more than 1,000 colonoscopies before the study.

Before study initiation, all participating endoscopists completed a brief familiarization session with the CADe system to standardize interpretation of visual alerts.

During AI-on weeks, a proprietary deep learning–based CADe system implemented on a graphics processing unit-equipped workstation was connected to the endoscopy video output through a video capture interface for real-time lesion detection. The platform processed the live video stream during the withdrawal phase and displayed a visual alert on a dedicated monitor positioned within the endoscopist's field of view when a suspected lesion was identified. Because the system was used as a deployed clinical tool rather than developed in-house, detailed information regarding internal model architecture, training dataset composition, and algorithm-level validation metrics was not accessible to the investigators.

It is important to note that the CADe system was utilized exclusively for lesion detection. When a lesion was detected, the endoscopist – not the AI system – performed optical characterization based on mucosal and vascular pattern assessment using standard endoscopic criteria and recorded the predicted histologic classification before lesion resection. Clinical decision-making, including whether to resect or biopsy a lesion, was left to the discretion of the endoscopist.

During AI-off weeks, the CADe system was disconnected, and no automated feedback was provided.

All detected lesions were removed or biopsied according to current clinical guidelines. Resected specimens were submitted for histopathological examination, which served as the reference standard. Pathologists were blinded to study allocation and to the endoscopists' optical predictions.

Withdrawal time was defined as the time from cecal intubation to scope withdrawal, excluding time spent on therapeutic interventions.

Outcomes

The primary outcome was the ADR, defined as the proportion of colonoscopies in which at least one histologically confirmed adenoma was detected.[9]

Secondary outcomes included:

- Polyp detection rate (PDR)
- Adenomas per colonoscopy (APC)
- Lesion size, morphology, and location (Paris classification)
- Withdrawal time
- Diagnostic accuracy of endoscopist-performed optical biopsy.

Optical biopsy performance was evaluated on a per-polyp basis during AI-on weeks. Endoscopists documented their optical diagnosis for resected polyps before histopathological evaluation. The analysis included all resected polyps during AI-on weeks for which an endoscopist's optical diagnosis had been prospectively recorded (n=428). Sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated with corresponding 95% confidence intervals (CIs) using histopathology as the reference standard.

Statistical Analysis

Normality of continuous variables was assessed using the Shapiro–Wilk test. Normally distributed variables are presented as mean ± standard deviation, whereas non-normally distributed variables are reported as median with interquartile range. Categorical variables are presented as counts and percentages.

Between-group comparisons were performed using the Student's t-test or the Mann–Whitney U test for continuous variables and the Chi-square test or Fisher's exact test for categorical variables, as appropriate.

To evaluate the independent association between AI assistance and adenoma detection, a multivariable logistic regression model was constructed adjusting for age, sex, colonoscopy indication, bowel preparation quality, and withdrawal time. Withdrawal time was included as a covariate because longer inspection time is an established predictor of adenoma detection, and we aimed to isolate the effect **o**f AI independent of minor procedural time variations. Adjusted odds ratios (ORs)

with 95% CIs were reported. Model calibration was assessed using the Hosmer–Lemeshow goodness-of-fit test.

All statistical tests were two-sided, and a p<0.05 was considered statistically significant. Statistical analyses were performed using the Statistical Package for the Social Sciences version 28.0 (IBM Corp., Armonk, NY, USA).

Sample Size Calculation

Based on our institution's baseline ADR of 13% observed in the year preceding the study, we aimed to detect an absolute increase of 6% points (13–19%), consistent with effect sizes reported in meta-analyses of CADe trials.

Assuming a two-sided alpha level of 0.05 and 80% power, approximately 1,000 patients per group were required. We enrolled 1,061 patients in each arm (total *n*=2,122), providing adequate statistical power for the primary analysis.

RESULTS

Study Population

A total of 2,122 consecutive colonoscopies were included in the analysis, with 1,061 procedures allocated to the AI-off group and 1,061 to the AI-on group. The study flow diagram is presented in Figure 1. Briefly, 2,345 colonoscopy procedures were assessed for eligibility; 223 were excluded due to inflammatory bowel disease (*n*=28), known CRC (*n*=17), hereditary syndromes (*n*=5), prior colorectal resection (*n*=41), emergency colonoscopy (*n*=33), or declined consent (*n*=99). No procedure was excluded based on bowel preparation quality. Baseline characteristics were comparable between groups (Table 1). The mean age was 58.4±12.1 years in the AI-off group and 57.9±11.8 years in the AI-on group (p=0.38). The proportion of male patients was 52.1% and 53.4%, respectively (p=0.56). Colonoscopy indications (screening, surveillance, or diagnostic) were similarly distributed between groups (p=0.72). Adequate bowel preparation (BBPS score ≥6) was achieved in 94.6% of procedures in the AI-off group and 95.2% in the AI-on group (p=0.49). Mean withdrawal time was 7.8±2.1 min in the AI-off group and 8.0 ± 2.3 min in the AI-on group (p=0.11).

Primary Outcome: ADR

AI assistance significantly improved adenoma detection. At least one histologically confirmed adenoma was detected in 138 of 1,061 colonoscopies (13.0%) in the AI-off group and in 202 of 1,061 colonoscopies (19.0%) in the AI-on group, corresponding to an absolute increase of 6.0% points (p<0.001) and a relative increase of 46%. The difference in ADR between groups is illustrated in Figure 2.

Secondary Outcomes

PDR was higher in the AI-on group than in the AI-off group (35.4% vs. 28.7%, p=0.002). Similarly, the mean number of APC

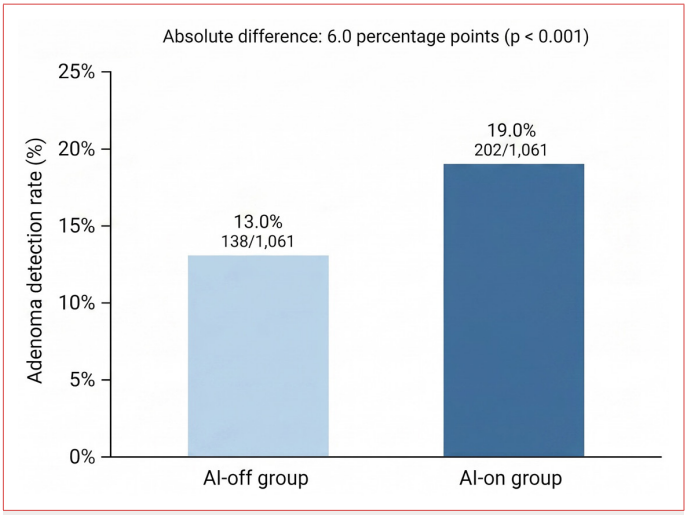

Figure 1. Study flow diagram. Flow diagram illustrating patient inclusion and allocation to study groups. A total of 2,345 colonoscopy procedures were assessed for eligibility between October 01, 2025, and February 01, 2026. Of these, 223 were excluded due to inflammatory bowel disease (*n*=28), known colorectal cancer (CRC) (*n*=17), hereditary CRC syndromes (*n*=5), prior colorectal resection (*n*=41), emergency colonoscopy (*n*=33), or declined consent (*n*=99). No procedure was excluded based on bowel preparation quality. The remaining 2,122 colonoscopies were allocated by weekly alternation: 1,061 to the AI-off group and 1,061 to the AI-on group

increased from 0.21±0.54 in the AI-off group to 0.32±0.71 in the AI-on group (p=0.004).

The increased detection rate was primarily driven by diminutive (≤5 mm) and non-polypoid lesions. The proportion of detected polyps measuring ≤5 mm increased from 58% in the AI-off group to 67% in the AI-on group (p=0.008). Non-polypoid lesions (Paris classification 0-II) were detected in 2.1% of AI-on colonoscopies compared with 1.1% in the AI-off group (p=0.02). Detailed lesion characteristics are presented in Table 2, and the distribution of polyp size categories across groups is illustrated in Figure 3.

Multivariable Analysis

After adjustment for age, sex, colonoscopy indication, bowel preparation quality, and withdrawal time, AI assistance remained independently associated with adenoma detection. AI-assisted colonoscopy was associated with higher odds of detecting at least one adenoma (adjusted OR 1.56; 95% CI 1.22–1.99; p<0.001). Older age, screening indication, and longer withdrawal time were also independently associated with adenoma detection. The complete regression model is presented in Table 3.

Table 1. Baseline characteristics of the study population

Variable	AI-off (*n*=1.061)	AI-on (*n*=1.061)	*p*
Age, years (mean±SD)	58.4±12.1	57.9±11.8	0.38
Male sex, *n* (%)	553 (52.1)	567 (53.4)	0.56
Indication for colonoscopy, *n* (%)			
Screening	412 (38.8)	426 (40.1)	0.72
Surveillance	265 (25.0)	251 (23.7)	
Diagnostic	384 (36.2)	384 (36.2)	
Adequate bowel preparation (BBPS ≥6), *n* (%)	1003 (94.6)	1011 (95.2)	0.49
Withdrawal time, minutes (mean±SD)	7.8±2.1	8.0±2.3	0.11

SD: Standard deviation, BBPS: Boston bowel preparation scale, AI: Artificial intelligence

Optical Biopsy Performance

During AI-on weeks, endoscopists were instructed to prospectively record optical diagnoses for resected polyps before histopathological evaluation. A total of 428 resected polyps with complete prospective optical diagnosis documentation were included in this sub-analysis. On a per-polyp basis, endoscopist-performed optical biopsy demonstrated a sensitivity of 88.4% (95% CI 83.6–92.1), specificity of 81.2% (95% CI 74.9–86.3), PPV of 84.7% (95% CI 80.1–88.5), and NPV of 85.4% (95% CI 80.8–89.1). Overall diagnostic accuracy was 85.1% (95% CI 81.4–88.3). For diminutive rectosigmoid polyps (≤5 mm), the NPV for adenomatous histology was 91.2% (95% CI 86.5–94.8). Table 4 summarizes the diagnostic performance of endoscopist-performed optical biopsy, and Figure 4 illustrates these estimates with corresponding CIs.

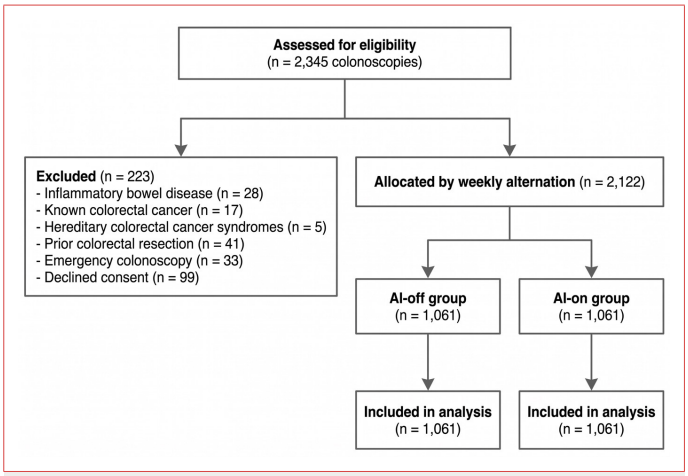

Figure 2. Comparison of adenoma detection rate between study groups. Adenoma detection rate (ADR) according to the study group. ADR was higher in the artificial intelligence (AI)-on group (19.0%, 202/1,061) compared with the AI-off group (13.0%, 138/1,061), corresponding to an absolute difference of 6.0% points (p<0.001)

DISCUSSION

In this prospective pragmatic comparative study, implementation of a deep learning–based CADe system in routine colonoscopy practice was associated with a significant increase in ADR. ADR improved from 13% to 19%, corresponding to an absolute increase of six percentage points and a relative increase of 46%. Importantly, this association remained significant after adjustment for potential confounders, with AI assistance independently associated with higher odds of adenoma detection.

The observed improvement in ADR is consistent with a growing body of randomized evidence. Multiple randomized

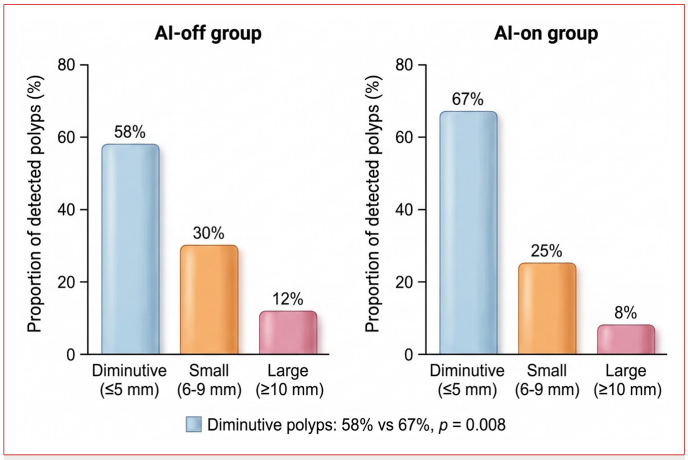

Figure 3. Distribution of polyp size categories in artificial intelligence (AI)-off and AI-on groups. Distribution of detected polyps according to size category. The proportion of diminutive polyps (≤5 mm) was higher in the AI-on group (67%) than in the AI-off group (58%) (p=0.008)

Table 2. Primary and secondary detection outcomes

Outcome	AI-off (*n*=1,061)	AI-on (*n*=1,061)	*p*
ADR, *n* (%)	138 (13.0)	202 (19.0)	<0.001
PDR, *n* (%)	305 (28.7)	376 (35.4)	0.002
APC, mean±SD	0.21±0.54	0.32±0.71	0.004
Diminutive polyps (≤5 mm), *n* (% of detected polyps)	177 (58)	252 (67)	0.008
Non-polypoid lesions (Paris 0-II), *n* (% of colonoscopies)	12 (1.1)	22 (2.1)	0.02
Withdrawal time, minutes (mean±SD)	7.8±2.1	8.0±2.3	0.11

ADR: Adenoma detection rate; PDR: Polyp detection rate; APC: Adenomas per colonoscopy; SD: Standard deviation; AI: Artificial intelligence.

Table 3. Multivariable logistic regression analysis for adenoma detection

Variable	Adjusted OR	95% CI	p
AI assistance (AI-on vs. AI-off)	1.56	1.22–1.99	<0.001
Age (per year increase)	1.02	1.01–1.03	0.01
Male sex	1.18	0.95–1.46	0.12
Screening indication (vs. diagnostic)	1.24	1.01–1.54	0.04
Adequate bowel preparation	1.31	0.90–1.89	0.15
Withdrawal time (per minute increase)	1.06	1.01–1.12	0.03

OR: Odds ratio; CI: Confidence interval; AI: Artificial intelligence.

controlled trials have demonstrated that CADe systems improve the detection of colorectal neoplasia (4–6). A comprehensive meta-analysis of randomized trials reported that AI-assisted colonoscopy significantly increases ADR (relative risk [RR] 1.19; 95% CI 1.14–1.25),[10] while another meta-analysis reported a similar magnitude of effect (RR 1.27; 95% CI 1.14–1.41).[11] Our findings extend this evidence by evaluating AI implementation in a pragmatic clinical setting using a non-randomized weekly alternation design rather than in tightly controlled experimental environments. The consistency between our results and those from randomized trials suggests that the benefits of CADe can be maintained in routine clinical practice.

The increase in detection observed in our cohort was primarily driven by diminutive (≤5 mm) and non-polypoid lesions. This pattern is consistent with previous studies showing that AI systems preferentially improve the detection of subtle lesions that are easily overlooked during conventional inspection.[3,6,12] Non-polypoid morphology has been associated with higher miss rates in tandem colonoscopy studies.[3] By facilitating the detection of these lesions, AI-assisted colonoscopy may contribute to reducing the risk of interval CRC. Although long-term oncologic outcomes cannot be determined from the present study, ADR is a validated surrogate marker for

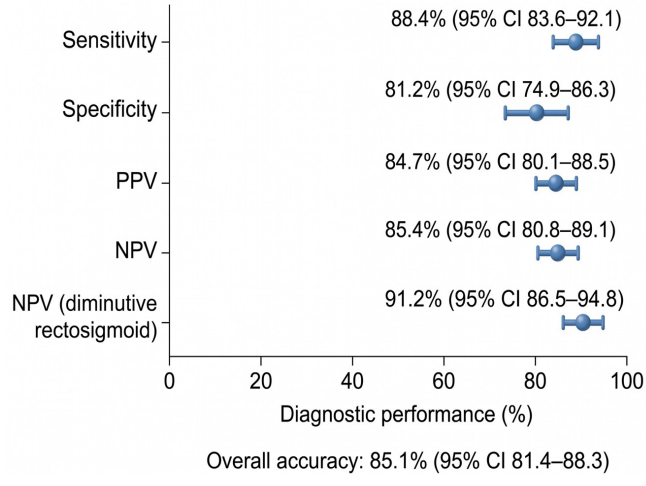

Figure 4. Diagnostic performance of endoscopist-performed optical biopsy. Forest plot showing diagnostic performance of endoscopist-performed optical biopsy, including sensitivity, specificity, positive predictive value, and negative predictive value, with corresponding 95% confidence intervals. Overall accuracy was 85.1% (95% confidence ınterval [CI] 81.4–88.3). For diminutive rectosigmoid polyps (≤5 mm), the negative predictive value was 91.2% (95% CI 86.5–94.8).

Table 4. Diagnostic performance of endoscopist-performed optical biopsy (per-polyp analysis)

Overall performance (n=428 polyps)		
Metric	**Estimate (%)**	**95% CI**
Sensitivity	88.4	83.6–92.1
Specificity	81.2	74.9–86.3
PPV	84.7	80.1–88.5
NPV	85.4	80.8–89.1
Overall accuracy	85.1	81.4–88.3
Subgroup: Diminutive rectosigmoid polyps (≤5 mm)		
Metric	**Estimate (%)**	**95 CI**
NPV	91.2	86.5–94.8

NPV: Negative predictive value; PPV: Positive predictive value; CI: Confidence interval.

CRC risk reduction.[2,13] Furthermore, a recent meta-analysis demonstrated that CADe significantly reduces adenoma miss rates (OR 0.58; 95% CI 0.45–0.74),[14] supporting the mechanistic basis of our findings.

Importantly, the implementation of AI assistance did not significantly prolong withdrawal time. This finding is consistent with prior randomized trials and pooled analyses indicating that CADe integration does not substantially increase procedure duration or disrupt endoscopic workflow.[15,16] In our study, withdrawal times in both groups remained within accepted quality standards, supporting the practical feasibility of AI-assisted colonoscopy in routine practice.

In addition to detection performance, we evaluated endoscopist-performed optical biopsy within an AI-assisted detection environment. Optical characterization demonstrated robust diagnostic performance, with a sensitivity of 88% and a specificity of 81%. For diminutive rectosigmoid polyps, the NPV exceeded 90%, meeting the preservation and ıncorporation of valuable endoscopic innovations threshold required for adopting a "leave-*in situ*" strategy.[7] Although our study was not designed to alter resection strategies, these findings suggest that endoscopist-performed optical diagnosis during AI-assisted colonoscopy may be feasible in routine clinical practice. Similar diagnostic performance has been reported in prospective multicenter studies evaluating real-time optical diagnosis of colorectal polyps.[17]

Structured diagnostic frameworks may further improve optical biopsy performance. Recent studies suggest that standardized approaches – such as the "3-s rule" for high-confidence optical diagnosis – can enhance diagnostic accuracy when combined with AI assistance.[18] In an *ex vivo* study including 35 endoscopists, AI-assisted optical characterization increased high-confidence diagnostic accuracy from 78.3% to 89.8%, with the largest improvements observed among less experienced endoscopists.[18] Although such structured decision rules were not implemented in our study, these findings may inform future strategies for improving optical diagnosis in clinical practice.

While controlled trials have demonstrated promising results for AI-assisted detection and optical diagnosis, real-world validation remains comparatively limited.[19,20] Soleymanjahi et al.[21] emphasized the importance of pragmatic evaluations beyond efficacy trials. Our findings contribute to this emerging evidence by demonstrating improved adenoma detection and stable diagnostic performance in a heterogeneous patient population representative of routine clinical practice. In addition, network meta-analyses indicate that improvements in adenoma detection appear consistent across different CADe platforms, suggesting that the observed benefit is not restricted to a specific device.[22] It should be noted that, as the present study employed a commercially available CADe system, detailed technical specifications regarding the underlying algorithm (e.g., architecture, training data) were not accessible to the investigators; this limitation should be considered when interpreting the findings.

Recent real-world investigations have also raised the possibility of unintended consequences associated with AI implementation. A retrospective study evaluating CAD integration reported that although ADR and PDR improved with AI-assisted colonoscopy, detection rates in non-AI procedures performed after AI introduction declined compared with the pre-implementation baseline.[23] This phenomenon may reflect changes in endoscopist attention or increased reliance on AI alerts. Although our study was not designed to evaluate such carryover effects, this observation highlights the importance of careful implementation strategies and ongoing quality monitoring.

Several limitations should be acknowledged. First, although the relative increase in ADR was substantial, the absolute increase was modest, consistent with pooled estimates reporting absolute improvements of approximately 5–10% depending on baseline ADR and study design.[24] Second, the non-randomized weekly alternation design, while pragmatic, may introduce temporal confounding despite the comparable case mix observed between groups; residual confounding cannot be entirely excluded. Third, the single-center design may limit generalizability to other settings with different patient populations or endoscopist experience levels. Fourth, long-term effects on interval CRC incidence cannot be evaluated in this study. Fifth, cost-effectiveness was not assessed, although

previous modeling studies suggest potential economic benefits of AI-assisted colonoscopy in screening programs.[25] Finally, as a pragmatic study using a commercial CADe platform, detailed technical parameters of the AI system were not available for analysis, which precludes algorithm-specific comparisons.

Despite these limitations, the study has several strengths. The pragmatic design reflects routine clinical workflow, participating endoscopists were experienced and evenly distributed between study groups, histopathology served as the reference standard, and multivariable analysis confirmed the independent association between AI assistance and improved adenoma detection. Moreover, the simultaneous evaluation of CADe performance and endoscopist-performed optical biopsy within the same cohort provides a comprehensive assessment of AI integration into routine colonoscopy practice.

Overall, these findings suggest that integration of AI-based detection and endoscopist-performed optical characterization into routine colonoscopy is feasible and may improve detection of colorectal neoplasia without compromising procedural efficiency.

CONCLUSION

In this prospective pragmatic comparative study of 2,122 colonoscopies, integration of real-time CADe into routine colonoscopy practice was associated with a significant increase in ADR from 13% to 19%. Endoscopist-performed optical characterization during AI-assisted colonoscopy demonstrated reliable diagnostic performance, with an NPV exceeding 90% for diminutive rectosigmoid polyps. These findings suggest that AI-assisted colonoscopy may be feasible in routine clinical practice, although further multicenter studies with longer follow-up are required to determine whether improved detection translates into sustained reductions in CRC incidence and mortality.

DECLARATIONS

Ethics Committee Approval: The study protocol was approved by the Kayseri City Hospital Non-Interventional Clinical Research Ethics Committee (Approval No: 585; September 26, 2025).

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

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: The authors declare that no artificial intelligence (AI)–assisted technologies (including large language models, chatbots, or image generators) were used in the generation of the scientific content of this manuscript.

Authorship Contributions: Concept – YO; Design – YO, NBM; Supervision – NMB; Data collection &/or processing – YO, FO, DT, OB; Analysis and/or interpretation – YO, NMB; Literature search – YO; Writing – YO; Critical review – NMB.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71:209–49.
2. Corley DA, Jensen CD, Marks AR, Zhao WK, Lee JK, Doubeni CA, et al. Adenoma detection rate and risk of colorectal cancer and death. N Engl J Med 2014;370:1298–306.
3. Zhao S, Wang S, Pan P, Xia T, Chang X, Yang X, et al. Magnitude, risk factors, and factors associated with adenoma miss rate of tandem colonoscopy: a systematic review and meta-analysis. Gastroenterology 2019;156:1661–74.e11.
4. Glissen Brown JR, Mansour NM, Wang P, Chuchuca MA, Minchenberg SB, Chandnani M, et al. Deep learning computer-aided polyp detection reduces adenoma miss rate: a United States multi-center randomized tandem colonoscopy study (CADeT-CS Trial). Clin Gastroenterol Hepatol 2022;20:1499–507.e4.
5. Repici A, Spadaccini M, Antonelli G, Correale L, Maselli R, Galtieri PA, et al. Artificial intelligence and colonoscopy experience: lessons from two randomised trials. Gut 2022;71:757–65.
6. Wallace MB, Sharma P, Bhandari P, East J, Antonelli G, Lorenzetti R, et al. Impact of artificial intelligence on miss rate of colorectal neoplasia. Gastroenterology 2022;163:295–304.e5.
7. Hassan C, Balsamo G, Lorenzetti R, Zullo A, Antonelli G. Artificial intelligence allows leaving-in-situ colorectal polyps. Clin Gastroenterol Hepatol 2022;20:2505–13.e4.
8. Lai EJ, Calderwood AH, Doros G, Fix OK, Jacobson BC. The Boston bowel preparation scale: a valid and reliable instrument for colonoscopy-oriented research. Gastrointest Endosc 2009;69:620–5.
9. East JE, Vleugels JL, Roelandt P, Bhandari P, Bisschops R, Dekker E, et al. Advanced endoscopic imaging: european society of gastrointestinal endoscopy (ESGE) technology review. Endoscopy 2016;48:1029–45.
10. Qutob IA, Soliman A, Trabelsi R, Elkholy MKA, Elgahamy AS, Abuismail M, et al. Improving colorectal cancer detection with ai-assisted colonoscopy: a systematic review and meta-analysis of 38 RCTs with GRADE assessment. J Gastrointest Cancer 2025;56:240.

11. Tan S, Zeng P, Liu S, Yang Y, Chen S, Zhang W, et al. Effectiveness of artificial intelligence-assisted colonoscopy in detecting and diagnosing colorectal tumors: a systematic review and network meta-analysis. Int J Colorectal Dis 2025;40:218.
12. Kaminski MF, Wieszczy P, Rupinski M, Wojciechowska U, Didkowska J, Kraszewska E, et al. Increased rate of adenoma detection associates with reduced risk of colorectal cancer and death. Gastroenterology 2017;153:98–105.
13. Lee JK, Jensen CD, Levin TR, Doubeni CA, Zauber AG, Chubak J, et al. Long-term risk of colorectal cancer and related death after adenoma removal in a large, community-based population. Gastroenterology 2020;158:884–94.e5.
14. Wang SY, Gao JC, Wu SD. Artificial intelligence for reducing missed detection of adenomas and polyps in colonoscopy: A systematic review and meta-analysis. World J Gastroenterol 2025;31:105753.
15. Wang P, Liu X, Berzin TM, Glissen Brown JR, Liu P, Zhou C, et al Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADe-DB trial): a double-blind randomised study. Lancet Gastroenterol Hepatol 2020;5:343–51.
16. Chang PW, Nguyen DD, Kong N, Wang D, Wang S, Ong J, et al. Impact of artificial intelligence–assisted colonoscopy on gastroenterology fellow performance: a pragmatic randomized controlled trial. Gastrointest Endosc 2026;103:1043–51.
17. Xu X, Ba L, Lin L, Song Y, Zhao C, Yao S, et al. Evaluation efficacy and accuracy of a real-time computer-aided polyp detection system during colonoscopy: a prospective, multicentric, randomized, parallel-controlled study trial. Surg Endosc 2025;39:7417–27.
18. Kang HY, Kang S, Chung GE, Baek DH, Lee HS, Park J, et al Structured integration of an artificial intelligence-based system for the optical diagnosis of colorectal polyps. Gut Liver 2026;20:86–96.
19. Mori Y, Kudo SE, Misawa M, Saito Y, Ikematsu H, Hotta K, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. Ann Intern Med 2018;169:357–66.
20. Djinbachian R, Dubé AJ, von Renteln D. Optical diagnosis of colorectal polyps: recent developments. Curr Treat Options Gastroenterol 2019;17:99–114.
21. Soleymanjahi S, Huebner J, Elmansy L, Rajashekar N, Lüdtke N, Paracha R, et al. Artificial intelligence-assisted colonoscopy for polyp detection : a systematic review and meta-analysis. Ann Intern Med 2024;177:1652–63.
22. Kumar S, Maheshwari M, Aleem S, Batool Z, Alsubaie N, Syed S, et al. Novel artificial intelligence systems in detecting adenomas in colonoscopy: a systemic review and network meta-analysis. Clin Transl Gastroenterol 2025;16:e00904.
23. Takasu A, Kogure H, Dai Z, Yamada Y, Nakayama M, Bechara R, et al. Impact of introducing artificial intelligence on colonoscopy: a retrospective study on potential benefits and drawbacks. J Gastroenterol Hepatol 2025;40:2258–66.
24. Lou S, Du F, Song W, Xia Y, Yue X, Yang D, et al. Artificial intelligence for colorectal neoplasia detection during colonoscopy: a systematic review and meta-analysis of randomized clinical trials. EClinicalMedicine 2023;66:102341.
25. Wong MC, Huang J, Lam TY, Lau LH, Chiu PW. The cost-effectiveness of ai-assisted colonoscopy as a primary or secondary screening test in a population-based colorectal cancer screening program: markov modeling-based cost effectiveness analysis. J Med Internet Res 2025;27:e67762.

ORIGINAL ARTICLE

European Archives of Medical Research

DOI: 10.14744/eamr.2026.60566
Eur Arch Med Res 2026;42(2):206–212

Premature Adrenarche Followed to Adolescence: PCOS Risk and Phenotype

Didem Helvacioglu, Aylin Tugba Canbaz, Busra Gurpinar Tosun, Tulay Guran

Department of Pediatric Endocrinology, Marmara University Faculty of Medicine, Istanbul, Türkiye

ABSTRACT

Objective: To evaluate the development of polycystic ovary syndrome (PCOS) during adolescence in girls with a history of premature adrenarche (PA) and to identify clinical, biochemical, and familial factors associated with subsequent PCOS.

Materials and Methods: This retrospective longitudinal cohort study included 88 girls diagnosed with PA in childhood who were followed until adolescence. PCOS was evaluated during adolescence in clinically indicated participants according to contemporary guideline-based diagnostic criteria.

Results: During follow-up, 15 of 88 girls developed PCOS (cumulative incidence, 17.0%). At PA diagnosis, age, birth weight SDS, and body mass index (BMI)-SDS were similar between the PCOS and non-PCOS groups, and baseline androgen measures did not differ (dehydroepiandrosterone sulfate [DHEA-S], p=0.35; DHEA-S SDS, p=0.80; total testosterone, p=0.87; Δ4-androstenedione, p=0.69). At adolescent/final evaluation, BMI-SDS and ΔBMI-SDS were comparable between the groups (p=0.84 and p=0.58, respectively). Girls who developed PCOS also had menstrual irregularity and clinical hyperandrogenism, with a median modified Ferriman-Gallwey score of 15 (IQR, 11–21). Maternal history of PCOS was more frequent among girls who developed PCOS than among those who did not (7/15 [46.7%] vs. 13/65 [20.0%], p=0.047; OR, 3.50). Total testosterone levels were significantly higher in the PCOS group (p=0.005), whereas DHEA-S and Δ4-androstenedione levels did not differ significantly between the groups.

Conclusion: In girls with PA, adrenal androgen levels at diagnosis did not predict subsequent PCOS. Familial susceptibility was associated with later PCOS development. At adolescent evaluation, total testosterone distinguished girls with PCOS from those without PCOS better than DHEA-S or Δ4-androstenedione. These findings support the continued follow-up of girls with PA through adolescence, with particular attention to familial predisposition, menstrual regularity, and clinical or biochemical hyperandrogenism.

Keywords: Adolescent, Adrenarche, Hyperandrogenism, Polycystic Ovary Syndrome, Risk Factors

Cite this article as: Helvacioglu D, Canbaz AT, Gurpinar Tosun B, Guran T. Premature Adrenarche Followed to Adolescence: PCOS Risk and Phenotype. Eur Arch Med Res 2026;42(2):206–212.

Address for correspondence: Didem Helvacioglu. Department of Pediatric Endocrinology, Marmara University Faculty of Medicine, Istanbul, Türkiye

E-mail: drdidemh@gmail.com **ORCID ID:** 0000-0001-8687-6326

Submitted: 08.04.2026 **Revised:** 26.04.2026 **Accepted:** 11.05.2026 **Available Online:** 03.06.2026

European Archives of Medical Research – Available online at www.eurarchmedres.org

OPEN ACCESS This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

INTRODUCTION

Premature adrenarche (PA) is characterized by early clinical signs of androgen action (pubic/axillary hair, acne, and body odor) before 8 years of age in girls and before 9 years of age in boys, after the exclusion of congenital adrenal hyperplasia, androgen-producing tumors, and other rare causes of androgen excess.[1] PA accounts for most isolated premature pubarche (PP) presentations and is usually considered a developmental variant but has been linked to later hyperandrogenic and metabolic risk in selected phenotypes.[1]

Polycystic ovary syndrome (PCOS) is among the most common endocrine disorders and carries a substantial reproductive and metabolic burden; guideline data estimate a global prevalence of ~10–13% and highlight persistent under-recognition.[2] In adolescents, diagnosis requires caution because physiological puberty may mimic PCOS; thus, recommendations require persistent menstrual irregularity (by years since menarche) and clinical or biochemical hyperandrogenism after the exclusion of mimicking disorders, while pelvic ultrasonography and AMH are not recommended for diagnosis in this age group.[3]

A biologically plausible link between PA and the later development of PCOS has long been proposed. Early follow-up studies suggested increased postpubertal hyperandrogenic ovarian phenotypes after premature pubarche.[4,5] However, more recent longitudinal cohorts have reported heterogeneous outcomes, with some showing limited ovarian dysfunction and others reporting a higher PCOS risk and highlighting maternal PCOS history and adiposity as correlates.[6,7] Accordingly, whether PA represents a benign variant or an early marker of future PCOS in specific subgroups remains clinically important. Therefore, we aimed to evaluate adolescent PCOS outcomes after PA and to identify clinical and biochemical predictors at presentation and during follow-up that may improve risk stratification.

MATERIALS AND METHODS

This retrospective longitudinal cohort study included girls diagnosed with premature adrenarche (PA) in childhood and followed through adolescence at a tertiary pediatric endocrinology center. At presentation, alternative causes of androgen excess were excluded, including classic/non-classic congenital adrenal hyperplasia, androgen-secreting tumors, Cushing syndrome, and central precocious puberty, using appropriate clinical, biochemical, and imaging assessments.[1,8]

Clinical history and follow-up data included age at symptom onset, birth weight, gestational age, maternal PCOS history, and maternal age at menarche. These data were obtained from medical records. Missing maternal history data were verified by telephone interview with the mother when possible.

Physical examination and anthropometric assessment were performed at 3–6-month intervals. Height and weight were measured using standardized methods; BMI was calculated as kg/m^2 and converted to age- and sex-specific SDS using Turkish reference data.[9] Auxologic and skeletal assessment included BMI-SDS change, bone age, and birth weight SDS. ΔBMI-SDS was defined as BMI-SDS at outcome evaluation minus BMI-SDS at PA diagnosis. Birth weight was expressed as SDS for gestational age; small for gestational age (SGA) was defined as birth weight <−2.0 SDS.[10]

Bone age (BA) and chronological age (CA) were recorded at PA diagnosis, and BA was assessed using the Greulich–Pyle method.[11]

Adolescent PCOS evaluation followed the 2023 International Evidence-based Guideline and adolescent-specific recommendations.[2,3] PCOS diagnosis required both ovulatory dysfunction, based on years-since-menarche criteria, and clinical and/or biochemical hyperandrogenism after the exclusion of mimicking disorders. Physical examination findings related to androgen action, including pubic/axillary hair, acne, body odor, and hirsutism during adolescence, were reviewed. Clinical hyperandrogenism was assessed by the presence of hirsutism and/or severe acne; hirsutism was quantified using the modified Ferriman–Gallwey (mFG) score.[12]

Laboratory assessment at PA diagnosis included serum DHEA-S, androstenedione, total testosterone, and 17-hydroxyprogesterone (17-OHP). At the adolescent/final evaluation, biochemical assessment for comparison between the PCOS and non-PCOS groups included total testosterone, DHEA-S, and androstenedione. DHEA-S was measured using an automated chemiluminescence immunoassay (ADVIA Centaur, Siemens Diagnostics, Erlangen, Germany). DHEA-S concentrations >1084 nmol/L (40 µg/dL) were considered indicative of adrenal activation, and age- and sex-adjusted DHEA-S SDS values were calculated.[13] Total testosterone and androstenedione were measured using automated immunoassays on a Beckman Coulter platform (Beckman Coulter, Brea, CA, USA).

Imaging and metabolic assessments were performed during the PCOS work-up when clinically indicated. Metabolic testing included fasting glucose, fasting insulin, HOMA-IR, and HbA1c when available. HOMA-IR was calculated according to the homeostasis model assessment described by Matthews et al.[14] using the following formula: fasting insulin (µU/mL)×fasting plasma glucose (mg/dL)/405. Pelvic ultrasonography was used to assess ovarian morphology and ovarian volume. Ovarian volumes were calculated using the ellipsoid formula: volume=length×width×thickness×0.523.[15]

Statistical Analysis

Statistical analyses were performed using GraphPad Prism version 5.0 (GraphPad Software, San Diego, CA, USA). Normality was assessed using the Shapiro–Wilk test. Continuous variables are presented as mean±SD or median (IQR), as appropriate, and categorical variables as n (%). Between-group comparisons were performed using Student's t-test for normally distributed data, the Mann–Whitney U test for non-normally distributed data, and the chi-square test or Fisher's exact test for categorical variables, as appropriate. Two-sided p values <0.05 were considered statistically significant.

Ethics

This study was conducted in accordance with the principles of the Declaration of Helsinki. The study protocol was approved by the Marmara University Institutional Ethics Committee (Approval No.: 09.2009.1020; Date: December 06, 2019), and written informed consent was obtained from the legal guardians.

RESULTS

A total of 88 girls were followed for PA diagnosed at a mean age of 6.6±1.1 years. The mean BMI SDS at presentation was 0.7±1.1. Most patients (81.9%) were born at term, and 8.0% were born SGA. The mean maternal age at menarche was 12.6±1.4 years; the corresponding mean age at menarche in the girls was 11.8±0.9 years. The mean BA/CA ratio was 1.0±0.2, and the mean DHEA-S concentration was 82.8±45.7 mcg/dL. The mean follow-up duration was 8.0±2.6 years, and PCOS was diagnosed in 15/88 girls (17.0%). Age at PA diagnosis was similar in the PCOS and non-PCOS groups (6.6±1.1 vs. 6.6±1.0 years; p=0.97). BMI-SDS at PA diagnosis did not differ between the PCOS and non-PCOS groups (0.9±1.2 vs. 0.6±1.1; p=0.47). Birth weight SDS did not differ between the groups (−0.57±1.10 vs. −0.54±0.96; p=0.93). SGA was observed in 2/15 (13.3%) girls in the PCOS group and 4/73 (5.5%) in the non-PCOS group (p=0.27). ΔBMI-SDS was similar between the groups (−0.08±1.3 vs. 0.09±1.0; p=0.58). In girls who developed PCOS, the mean age at menarche was 12.0±0.6 years, and the mean time from menarche to adolescent/final evaluation was approximately 5 years. Menarcheal age did not differ between girls who later developed PCOS and those who did not (p=0.70). At adolescent/final evaluation, age was 17.0±2.0 years in the PCOS group and 15.9±1.8 years in the non-PCOS group (p=0.057), while BMI-SDS was similar between the groups (0.70±2.3 vs. 0.8±1.6; p=0.84). At PCOS evaluation, all girls in the PCOS group reported menstrual irregularity, and clinical hyperandrogenism was evident, with a median Ferriman–Gallwey score of 15 (IQR, 11–21). Maternal history of PCOS was more frequent in girls who developed PCOS than in those who did not (7/15 [46.7%] vs. 13/65 [20.0%]; p=0.047; OR, 3.50; 95% CI, 1.07–11.42).

Maternal age at menarche did not differ between the PCOS and non-PCOS groups (13.0±0.8 vs. 12.6±1.4 years; p=0.40). In the non-PCOS group, daughters' age at menarche did not differ from maternal age at menarche (p=0.72). In contrast, in the PCOS group, daughters reached menarche significantly earlier than their mothers (p=0.001).

At PA diagnosis, neither DHEA-S concentration nor DHEA-S SDS differed between girls who later developed PCOS and those who did not (DHEA-S: 72.3±47.1 vs. 84.9±45.5, p=0.35; DHEA-S-SDS: 0.9±2.5 vs. 1.0±2.0, p=0.80); similarly, total testosterone and Δ4-androstenedione were comparable between the groups (p=0.87 and p=0.69, respectively) (Table 1).

Consistent with this clinical hyperandrogenic phenotype, total testosterone was significantly higher in the PCOS group than in the non-PCOS group (60.6±11.3 vs. 39.0±17.6, p=0.005), whereas Δ4-androstenedione (3.0±0.7 vs. 2.6±0.8, p=0.17) and DHEA-S (252±112 vs. 219±97.1, p=0.26) did not differ significantly between the groups. In the PCOS group, mean LH and FSH levels were 11.7±3.8 and 6.1±1.9, respectively. Metabolic assessment revealed a mean HOMA-IR of 3.7±2.2 and a mean HbA1c of 5.4±0.2%. Pelvic ultrasonography demonstrated mean ovarian volumes of 12.0±8.5 mL and 12.3±10.5 mL for the right and left ovaries, respectively, while peripheral follicular distribution was present in 40% of cases.

DISCUSSION

In this longitudinal cohort of girls with PA, 16.9% developed PCOS during adolescent follow-up. The main findings were that baseline clinical and biochemical characteristics at PA diagnosis, including age at presentation, birth weight, BMI-SDS, DHEA-S, DHEA-S SDS, total testosterone, and Δ4-androstenedione, did not differ between girls who subsequently developed PCOS and those who did not. In contrast, maternal history of PCOS was more frequent among girls who developed PCOS, and total testosterone at adolescent/final evaluation was the biochemical parameter that best differentiated PCOS status.

In the general adolescent population, the prevalence of PCOS is substantially lower than that observed after premature adrenarche, with recent meta-analytic data estimating a pooled prevalence of 6.3% using guideline-consistent adolescent criteria and 9.8% using the original Rotterdam criteria.[16] Our study found that 16.9% of girls with a history of PA developed PCOS during follow-up. This estimate falls within the range reported in longitudinal studies from Spain (15–40%).[17,18] Taken together, these findings support the view that premature adrenarche is associated with an increased risk of subsequent PCOS. Rather than representing only a benign variation in pubertal timing, PA may serve as an early clinical marker of susceptibility to later hyperandrogenism and PCOS development.

Table 1. Baseline and adolescent/final characteristics according to subsequent PCOS outcome

Variable	PCOS	Non-PCOS	p
Birth weight SDS	−0.57±1.10	−0.54±0.96	0.93
At PA diagnosis			
Age at PA diagnosis, years	6.6±1.1	6.6±1.0	0.97
BMI SDS	0.9±1.2	0.6±1.1	0.47
DHEA-S, mcg/dL	72.3±47.1	84.9±45.5	0.35
DHEA-S SDS	0.9±2.5	1.0±2.0	0.80
Total testosterone, ng/ml	0.09±0.06	0.09±0.08	0.87
Δ4-androstenedione, µg/L	0.6±0.4	0.6±0.4	0.69
17-hydroxyprogesterone (17-OHP), µg/L	1.1±0.9	0.8±0.5	0.10
At adolescent/final evaluation			
Age at adolescent/final evaluation, years	17.0±2.0	15.9±1.8	0.057
Age at menarche, years	12.0±0.6	11.7±1.0	0.70
BMI SDS	0.70±2.3	0.8±1.6	0.84
ΔBMI SDS	−0.08±1.3	0.09±1.0	0.58
Total testosterone, ng/dl	60.6±11.3	39.0±17.6	**0.005**
Δ4-androstenedione, µg/L	3.0±0.7	2.6±0.8	0.17
DHEA-S, mcg/dL	252±112	219±97.1	0.26

PCOS: Polycystic Ovary Syndrome; PA: Premature Adrenarche; SDS: Standard Deviation Score; BMI: Body Mass Index; DHEA-S: Dehydroepiandrosterone Sulfate.

In our cohort of girls with premature adrenarche, birth weight SDS did not differ between those who developed PCOS and those who did not (p=0.93), and although SGA was numerically more frequent in the PCOS group than in the non-PCOS group, this pattern was not statistically robust. These findings are consistent with the heterogeneous literature on prenatal growth and later PCOS risk. A prospective birth cohort study reported that women born SGA had approximately double the prevalence of PCOS compared with AGA peers (adjusted RR, 2.44; 95% CI, 1.39–4.28), supporting a developmental contribution of fetal growth restraint.[19]

However, more recent long-term follow-up data in premature adrenarche did not confirm SGA as an independent predictor (univariate OR, 0.94; p=0.92), with baseline BMI z-score and maternal PCOS history emerging as more relevant predictors.[7]

In our cohort, a maternal history of PCOS was present in 20/80 participants (25.0%) with available data and was more frequent among girls who developed PCOS than among those who did not (7/15 [46.7%] vs. 13/65 [20.0%]; OR, 3.50). This finding is consistent with the established familial aggregation of PCOS in first-degree relatives.[20] Maternal PCOS history has also been highlighted in PA/PP cohorts. Olgun et al.[21] reported maternal PCOS in ~15% of girls with premature pubarche. Efthymiadou et al.[22] found maternal PCOS in 33/89 (37.1%) girls with premature adrenarche. Differences across studies may reflect variation in the definition of family history, under-recognition of PCOS in mothers, and differences in outcome ascertainment.

Moreover, in the long-term follow-up of girls with premature adrenarche, Saroufim et al.[7] identified maternal PCOS history as the strongest risk factor for subsequent self-reported PCOS (OR, 4.1). Differences across studies likely reflect variation in the definition of family history, under-recognition of PCOS in mothers, and outcome ascertainment.

At the adolescent/final evaluation (mean age, 17.0±2.0 years in the PCOS group and 15.9±1.8 years in the non-PCOS group; p=0.057), the PCOS work-up in the PCOS group was performed approximately 5 years after menarche (mean age at menarche, 12.0±0.6 years), reducing the likelihood of misclassification due to normal pubertal maturation. With respect to pubertal tempo, age at menarche did not differ by later PCOS status in our cohort. While Guarnotta et al.[23] reported earlier menarche in girls with precocious pubarche who later developed PCOS (p=0.015). Menarche timing appears to be an inconsistent standalone predictor of PCOS risk and may be shaped by genetic background and adiposity.[24]

Maternal age at menarche was similar between girls who developed PCOS and those who did not (13.0±0.8 vs. 12.6±1.4 years; p=0.40). Although maternal PCOS has been associated with earlier pubertal development in daughters, this effect appears to be driven predominantly by adrenarche-related milestones rather than menarche.[25] In our cohort, an intergenerational shift toward earlier menarche was observed only among daughters who later developed PCOS. Nevertheless, Mendelian randomization data suggest that the inverse association between age at menarche and PCOS risk is modest, limiting the usefulness of menarcheal timing as a clinical discriminator.[26]

Girls diagnosed with PCOS in our cohort had a median mFG score of 15 (11–21), which is clearly above the recommended ethnicity-specific threshold (4–6) for detecting hirsutism.[2,3] This indicates that hirsutism was a major clinical feature at PCOS evaluation. As self-treatment is common and may reduce visible terminal hair growth, the recorded mFG values may underestimate the true severity of hirsutism.[2,3]

The laboratory findings showed that adrenal androgen levels at PA diagnosis had limited prognostic value. DHEA-S and DHEA-S SDS measured at the time of PA diagnosis did not differ between girls who later developed PCOS and those who did not (p=0.35 and p=0.80, respectively). Similarly, baseline total testosterone and Δ4-androstenedione were comparable between the groups. This aligns closely with Guarnotta et al.,[23] who evaluated girls with precocious pubarche both at diagnosis and again 10 years after menarche and reported similar DHEA-S values in those who later developed PCOS versus those who did not at both time points.

Likewise, in a long-term PA follow-up cohort, baseline androgen levels were not associated with later self-reported PCOS.[7] Taken together, these data suggest that the degree of adrenal androgen excess at PA presentation alone is insufficient to predict later PCOS.

In contrast, the adolescent biochemical profile was more informative. Importantly, most adolescents reach adult-range androgen concentrations by 12–15 years of age.[2] This supports the interpretability of biochemical hyperandrogenism assessed in our cohort at these ages. Total testosterone was significantly higher in the PCOS group, whereas DHEA-S and Δ4-androstenedione did not differ significantly between the groups. This pattern is clinically plausible and consistent with contemporary international recommendations, which prioritize total and free testosterone as the main biochemical markers of hyperandrogenism in adolescent PCOS, while DHEA-S and androstenedione are considered secondary markers with lower specificity.[2,3] Our findings are also consistent with longitudinal PA/PP literature suggesting that adrenal androgen markers, particularly DHEA-S, may reflect adrenarchal activation but do not reliably distinguish later PCOS phenotype on their own.[6,23] Accordingly, biochemical assessment in girls with a history of premature adrenarche should be interpreted together with menstrual pattern and persistent clinical hyperandrogenism rather than isolated adrenal androgen values.[2,3]

The hormonal profile in the PCOS group, characterized by higher LH relative to FSH (mean LH/FSH ratio ≈1.9), is consistent with the relative LH predominance described in PCOS. This pattern is thought to reflect altered hypothalamic-pituitary-ovarian regulation, with increased GnRH pulse frequency favoring LH secretion, thereby enhancing ovarian theca-cell androgen production and contributing to follicular arrest. However, as LH and the LH/FSH ratio are not recommended diagnostic criteria in adolescents, these findings should be interpreted as supportive of the underlying neuroendocrine phenotype rather than diagnostically specific.[27,28] The mean HOMA-IR of 3.7±2.2 in the PCOS group is consistent with increased insulin resistance in late/postpubertal adolescence, although interpretation should remain cautious given the known effects of pubertal stage and adiposity on HOMA-IR and the absence of a universally accepted adolescent threshold. Published pediatric cutoffs have ranged from 3.82 in pubertal girls to 2.91 in postpubertal adolescents.[29,30]

Hyperinsulinemia is pathophysiologically relevant in PCOS, as it may enhance ovarian androgen production and suppress hepatic SHBG synthesis, thereby aggravating the hyperandrogenic milieu; accordingly, the normal mean HbA1c in our cohort does not exclude early metabolic dysfunction.[2]

Pelvic ultrasonographic findings were described in girls with PCOS; however, ovarian morphology is not recommended as a diagnostic criterion for adolescent PCOS, particularly within 8 years after menarche.[2,3] Therefore, ultrasonographic findings in this cohort should be considered supportive descriptive data rather than diagnostic evidence.

This study has certain limitations. First, its retrospective single-center design may have led to missing data and variability in the timing of adolescent/final assessments. Second, adolescent PCOS evaluation was performed in clinically indicated participants rather than systematically in the entire cohort. Third, maternal PCOS history was unavailable for a subset of participants and was partly verified by telephone interview. Finally, metabolic parameters such as HOMA-IR and HbA1c were not systematically available for the non-PCOS group, limiting direct comparison of metabolic profiles between the groups.

Nonetheless, the long-term follow-up and the application of contemporary adolescent-specific diagnostic criteria provide clinically meaningful insight into the relationship between premature adrenarche and the later development of PCOS.

CONCLUSION

Our findings indicate that premature adrenarche is not uniformly benign and may represent an early clinical marker of the later development of PCOS in a susceptible subgroup. Early adrenal androgen levels at presentation were not predictive, whereas maternal PCOS history emerged as a more informative risk marker. These data support a longitudinal surveillance approach in girls with PA that extends beyond childhood and focuses on familial background, menstrual pattern, and evolving hyperandrogenic features during adolescence.

DECLARATIONS

Ethics Committee Approval: The study protocol was approved by Marmara University Institutional Ethics Committee (Approval No: 09.2009.1020; Date: December 06, 2019).

Informed Consent: Written informed consent was obtained from the legal guardians of the participants.

Conflict of Interest: The authors declare that there is no conflict of interest.

Funding: The authors received no financial support for the research and/or authorship of this article.

Use of AI for Writing Assistance: The authors declare that no artificial intelligence (AI)-assisted technologies were used in the generation of the scientific content, analysis, interpretation, or writing of this manuscript.

Authorship Contributions: Concept – DH, ATC, BGT, TG; Design – DH, ATC, BGT, TG; Supervision – DH, ATC, BGT, TG; Fundings – DH, ATC, BGT, TG; Materials – DH, ATC; Data collection &/or processing – DH, ATC; Analysis and/or interpretation – DH, BGT; Literature search – DH, TG; Writing – DH, TG; Critical review – DH, TG.

Peer-review: Externally peer-reviewed.

REFERENCES

1. Rosenfield RL. Normal and premature adrenarche. Endocr Rev 2021;42:783–814.
2. Teede HJ, Tay CT, Laven JJE, Dokras A, Moran LJ, Piltonen TT, et al. Recommendations from the 2023 international evidence-based guideline for the assessment and management of polycystic ovary syndrome. Eur J Endocrinol 2023;189:G43–64.
3. Peña AS, Witchel SF, Boivin J, Burgert TS, Ee C, Hoeger KM, et al. International evidence-based recommendations for polycystic ovary syndrome in adolescents. BMC Med 2025;23:151.
4. Ibañez L, Potau N, Virdis R, Zampolli M, Terzi C, Gussinyé M, et al. Postpubertal outcome in girls diagnosed of premature pubarche during childhood: increased frequency of functional ovarian hyperandrogenism. J Clin Endocrinol Metab 1993;76:1599–603.
5. Ibáñez L, Valls C, Potau N, Marcos MV, de Zegher F. Polycystic ovary syndrome after precocious pubarche: ontogeny of the low-birthweight effect. Clin Endocrinol (Oxf) 2001;55:667–72.
6. Tennilä J, Jääskeläinen J, Utriainen P, Voutilainen R, Häkkinen M, Auriola S, et al. PCOS features and steroid profiles among young adult women with a history of premature adrenarche. J Clin Endocrinol Metab 2021;106:e3335–45.
7. Saroufim R, Kurman T, Eugster EA. Long-term follow-up and risk of PCOS in indiana girls with a history of premature adrenarche-a single center experience. Children (Basel) 2025;12:1609.
8. Temeck JW, Pang SY, Nelson C, New MI. Genetic defects of steroidogenesis in premature pubarche. J Clin Endocrinol Metab 1987;64:609–17.
9. Bundak R, Furman A, Gunoz H, Darendeliler F, Bas F, Neyzi O. Body mass index references for Turkish children. Acta Paediatr 2006;95:194–8.
10. Battaglia FC, Lubchenco LO. A practical classification of newborn infants by weight and gestational age. J Pediatr 1967;71:159–63.
11. Greulich WW, Pyle SI. Radiographic atlas of skeletal development of the hand and wrist. 2nd ed. Stanford (CA): Stanford University Press; 1959.
12. Hatch R, Rosenfield RL, Kim MH, Tredway D. Hirsutism: implications, etiology, and management. Am J Obstet Gynecol 1981;140:815–30.
13. Guran T, Firat I, Yildiz F, Kaplan Bulut I, Dogru M, Bereket A. Reference values for serum dehydroepiandrosterone-sulphate in healthy children and adolescents with emphasis on the age of adrenarche and pubarche. Clin Endocrinol (Oxf) 2015;82:712–8.
14. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia 1985;28:412–9.
15. Caprio MG, Di Serafino M, De Feo A, Guerriero E, Perillo T, Barbuto L, et al. Ultrasonographic and multimodal

imaging of pediatric genital female diseases. J Ultrasound 2019;22:273–89.

16. Neven ACH, Forslund M, Ranashinha S, Mousa A, Tay CT, Peña A, et al. Prevalence and accurate diagnosis of polycystic ovary syndrome in adolescents across world regions: a systematic review and meta-analysis. Eur J Endocrinol 2024;191:S15–27.
17. Ibáñez L, Dimartino-Nardi J, Potau N, Saenger P. Premature adrenarche--normal variant or forerunner of adult disease? Endocr Rev 2000;21:671–96.
18. Utriainen P, Laakso S, Liimatta J, Jääskeläinen J, Voutilainen R. Premature adrenarche--a common condition with variable presentation. Horm Res Paediatr 2015;83:221–31.
19. Melo AS, Vieira CS, Barbieri MA, Rosa-E-Silva AC, Silva AA, Cardoso VC, et al. High prevalence of polycystic ovary syndrome in women born small for gestational age. Hum Reprod 2010;25:2124–31.
20. Kahsar-Miller MD, Nixon C, Boots LR, Go RC, Azziz R. Prevalence of polycystic ovary syndrome (PCOS) in first-degree relatives of patients with PCOS. Fertil Steril 2001;75:53–8.
21. Olgun EG, Cetin SK, Siklar Z, Aycan Z, Ozsu E, Ceran A, et al. Investigation of early puberty prevalence and time of addition thelarche to pubarche in girls with premature pubarche: two-year follow-up results. Clin Pediatr Endocrinol 2022;31:25–32.
22. Efthymiadou A, Bogiatzidou M, Kritikou D, Chrysis D. Anti-müllerian hormone in girls with premature adrenarche: the impact of polycystic ovary syndrome history in their mothers. J Pediatr 2019;205:190–4.
23. Guarnotta V, Lucchese S, Mineo MI, Mangione D, Venezia R, Almasio PL, et al. Predictive factors of polycystic ovary syndrome in girls with precocious pubarche. Endocr Connect 2021;10:796–804.
24. Carroll J, Saxena R, Welt CK. Environmental and genetic factors influence age at menarche in women with polycystic ovary syndrome. J Pediatr Endocrinol Metab 2012;25:459–66.
25. Lunddorf LLH, Arendt LH, Ernst A, Brix N, Knudsen UB, Olsen J, et al. Maternal polycystic ovarian syndrome and pubertal development in daughters and sons: a population-based cohort study. Hum Reprod 2022;37:2623–34.
26. Ma Y, Cai J, Liu LW, Hou W, Wei Z, Wang Y, et al. Age at menarche and polycystic ovary syndrome: A Mendelian randomization study. Int J Gynaecol Obstet 2023;162:1050–6.
27. Dong J, Rees DA. Polycystic ovary syndrome: pathophysiology and therapeutic opportunities. BMJ Med 2023;2:e000548.
28. Deans R. Polycystic ovary syndrome in adolescence. Med Sci (Basel) 2019;7:101.
29. Kurtoğlu S, Hatipoğlu N, Mazıcıoğlu M, Kendirici M, Keskin M, Kondolot M. Insulin resistance in obese children and adolescents: HOMA-IR cut-off levels in the prepubertal and pubertal periods. J Clin Res Pediatr Endocrinol 2010;2:100–6.
30. de Cassia da Silva C, Zambon MP, Vasques ACJ, Camilo DF, de Góes Monteiro Antonio MÂR, Geloneze B. The threshold value for identifying insulin resistance (HOMA-IR) in an admixed adolescent population: A hyperglycemic clamp validated study. Arch Endocrinol Metab 2023;67:119–25.