

CYPRUS

JOURNAL OF MEDICAL SCIENCES

Indexed in the Web of Science

Volume: **11** Issue: **5** October 2026

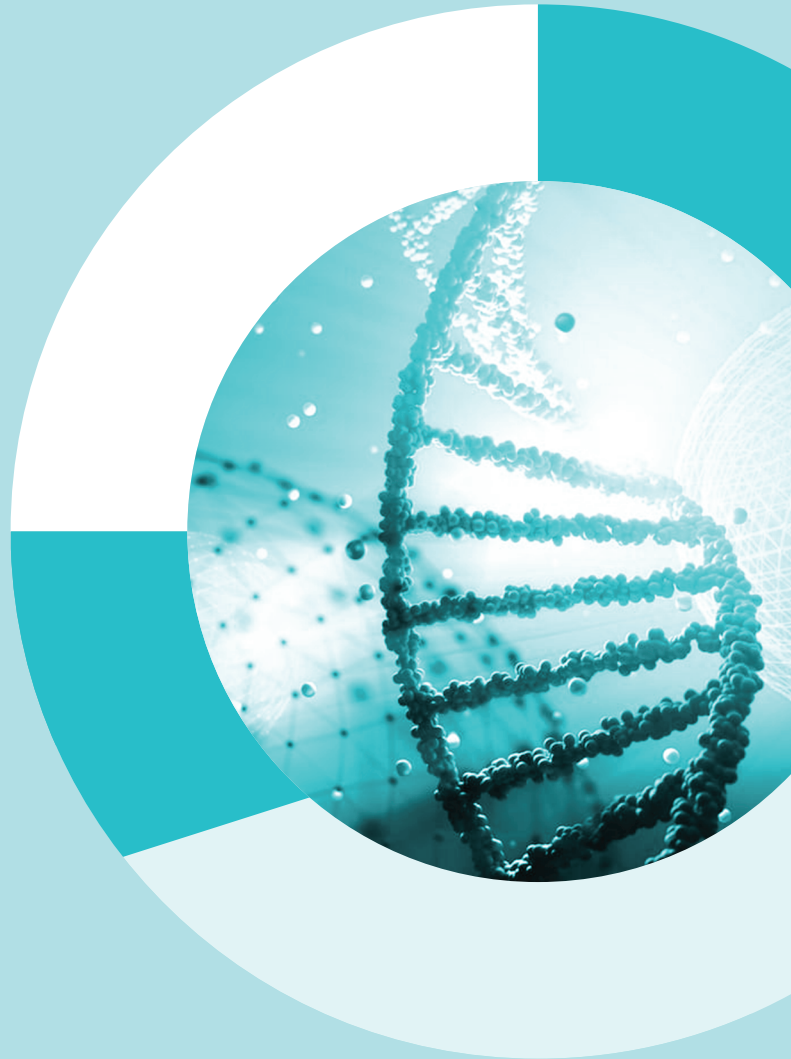


REVIEWS

- **Migration Related Tuberculosis**
Baha and Canbay.; Nicosia, Cyprus

RESEARCH ARTICLES

- **Depressive Symptoms in Children**
Cengiz et al.; Erzurum, Ağrı, Türkiye
- **Vaginal Cuff Closure Techniques in Total Laparoscopic Hysterectomy**
Peker et al.; Manisa, İzmir, Türkiye
- **PCOS Detection Using Gradient Boosting and XAI**
Karabulut Doğan et al.; Kayseri, Malatya, İstanbul, Türkiye
- **LLM Responses to Retinal Surgery Questions**
Duman et al.; Erbil, Iraq; Diyarbakır, İstanbul, Türkiye
- **Anterior Tarsal Tunnel Syndrome**
Azman-İste and Özbek.; İstanbul, Türkiye
- **Bibliometrics of TMJ Arthrocentesis**
İşleyen and Aydemir.; Burdur, Türkiye
- **Self-Management and Adjustment in Celiac Disease**
Bodur and Tüzer.; Ankara, Türkiye
- **Tobacco Control Compliance in Nicosia Hospitals**
Aşut et al.; Nicosia, Cyprus
- **MRI Signals and Breast Cancer Biomarkers**
Balyemez et al.; Nicosia, Cyprus; Iowa, United States; İstanbul, Türkiye
- **Screen Use and Speech-Language Disorders**
Ay et al.; Karabük, İstanbul, Ankara, Türkiye
- **Oocyte Donation Patient Profile in Northern Cyprus**
Burcu Özbakır.; Nicosia, Cyprus
- **Patient-Reported Outcomes after Cataract Surgery**
Keskinel and Dürüst Sakallı.; Nicosia, Famagusta, Cyprus



CYPRUS

JOURNAL OF MEDICAL SCIENCES

Indexed in Web of Science

Volume: 11 | Issue: 5 | October 2026

EDITORIAL BOARD

Editor-in-Chief

● Sonuç Büyük

Department of Pathology, Dr. Burhan Nalbantoğlu State Hospital, Nicosia, Cyprus

sonucbuyuk@outlook.com

<https://ease.org.uk/member-profile/sonuc-buyuk/>

Associate Editors

● Amber Eker Bakkaloğlu

Department of Neurology, Eastern Mediterranean University, Dr.

Fazıl Küçük Faculty of Medicine, Famagusta, Cyprus

amber.eker@emu.edu.tr

● Ayşe Baha

Department of Chest Diseases, Dr. Akçiçek State Hospital; Girne

American University Faculty of Medicine, Kyrenia, Cyprus

dr_aysedemir@hotmail.com

● Ayşe Ülgen

Department of Biostatistics, Girne American University Faculty

of Medicine, Kyrenia, Cyprus

ayseulgen1@gmail.com.tr

● Cemal Gürkan

Turkish Cypriot DNA Laboratory, Nicosia, Cyprus

Eastern Mediterranean University, Dr. Fazıl Küçük Faculty of

Medicine, Famagusta, Cyprus

cemal.gurkan@gmail.com

● Cenk Conkbayır

Department of Cardiology, Dr. Burhan Nalbantoğlu State

Hospital, Nicosia, Cyprus

cenkconk@hotmail.com

● Emil Mammadov

Department of Pediatric Surgery, Near East University Faculty of

Medicine, Nicosia, Cyprus

emil.mammadov@neu.edu.tr

● Erol Dülger

Vip Health Clinic, Nicosia, Cyprus

drerold@yahoo.com

● İzgen Karakaya

Department of Restorative Dentistry, European University of

Lefke, Faculty of Dentistry, Lefke, North Cyprus

izgen96h@gmail.com

● Mehtap Tınazlı

Head of the Department of Internal Medicine, Near East

University Faculty of Medicine Hospital, Nicosia, Cyprus

mehtap.tinazli@neu.edu.tr, mehtap.canbaz@hotmail.com

● Mümtaz Güran

Department of Medical Microbiology, Eastern Mediterranean

University, Dr. Fazıl Küçük Faculty of Medicine, Famagusta,

Cyprus

mumtazguran@gmail.com

● Nilüfer Güzoğlu

Department of Neonatology, Eastern Mediterranean University,

Dr. Fazıl Küçük Faculty of Medicine, Famagusta, Cyprus

nilufer.guzoglu@emu.edu.tr

● Özüm Buğdaycı

Department of Radiology, Cyprus International University

Faculty of Medicine, Nicosia, Cyprus

ozum.bugdayci@gmail.com

CYPRUS

JOURNAL OF MEDICAL SCIENCES

Indexed in Web of Science

Volume: 11 | Issue: 5 | October 2026

EDITORIAL BOARD

🔗 **Pınar Tunçbilek Özmanevra**

Department of Otorhinolaryngology - Head and Neck Surgery,
PrimeMed Clinic, Kyrenia, Cyprus
pinartuncbilek@gmail.com

🔗 **Ramadan Özmanevra**

Department of Orthopaedics and Traumatology, Cyprus
International University Faculty of Medicine, Nicosia, Cyprus
rozmanevra@gmail.com

🔗 **Gülten Sucu Dağ**

Department of Nursing, Eastern Mediterranean University
Faculty of Health Sciences, Famagusta, Cyprus
sucugulten@gmail.com

🔗 **Cenk Serhan Özverel**

Department of Basic Medical Sciences, Near East University
Faculty of Dentistry; DESAM Research Institute, Near East
University, Nicosia, North Cyprus
cenkserhan.ozverel@neu.edu.tr

🔗 **Safir Ullah Khan**

Moore's Cancer Center, University of California, San Diego,
United States
sak049@health.ucsd.edu

🔗 **İlker Etikan**

Department of Biostatistics, Near East University Faculty of
Medicine, Nicosia, Cyprus
ietikan@gmail.com

🔗 **Nail Bulakbaşı**

Department of Radiology, Dr. Suat Günsel University of Kyrenia
Hospital, Kyrenia, Cyprus
nbulakbasi@yahoo.com

🔗 **Özen Aşut**

Department of Public Health, Near East University Faculty of
Medicine, Nicosia, Cyprus
ozen.asut@neu.edu.tr

🔗 **Ceren M. Dönmezer Birand**

Department of Oral, Dental, and Maxillofacial Surgery, Near East
University Faculty of Dentistry, Nicosia, Cyprus
cerendonmezerbirand@gmail.com

🔗 **Müjgan Fırıncioğluları**

Department of Dentomaxillofacial Radiology, Cyprus
International University Faculty of Dentistry, Nicosia, Cyprus
mujganfirincioglulari@gmail.com

CYPRUS

JOURNAL OF MEDICAL SCIENCES

Indexed in Web of Science

Volume: 11 | Issue: 5 | October 2026

EDITORIAL BOARD

Section Editors

● Ahmet Özant

Private Clinic of Orthodontics, Nicosia, Cyprus
ozantahmet@gmail.com

● Ali Cenk Özay

Department of Obstetrics and Gynaecology, Near East University
Faculty of Medicine, Nicosia, Cyprus
drcenkozay@yahoo.com

● Ceyhan Dalkan

Department of Pediatrics, Division of Neonatology, Near East
University Faculty of Medicine, Nicosia, Cyprus
dalkanc@yahoo.com

● Ersan Berksel

Cyprus Science University Faculty of Health Sciences, Kyrenia,
Cyprus
ersanberksel@su.edu.tr

● Eşref Çelik

Department of Medical and Clinical Microbiology, Near East
University Faculty of Medicine, Nicosia, Cyprus
esref.celik@neu.edu.tr

● Gökçe Savtekin

Department of Oral and Maxillofacial Surgery, University of City
Island Faculty of Dentistry, Famagusta, Cyprus
gokcesavtekin@gmail.com

● Hülya Efetürk

Department of Nuclear Medicine, Near East University Faculty
of Medicine, Nicosia, Cyprus
drhulyaefeturk@gmail.com

● Hüseyin Kaya Sür

Department of Infectious Diseases and Clinical Microbiology,
Near East University Faculty of Medicine, Nicosia, Cyprus
kaya.suer@neu.edu.tr

● Necdet Özçay

Department of General Surgery, University of Health Sciences
Türkiye, Gülhane Faculty of Medicine, Ankara, Türkiye
necdetozcay@gmail.com

● Nedim Sezgin İlgi

Department of Anatomy, Near East University Faculty of
Medicine, Nicosia, Cyprus
sezgin.ilgi@neu.edu.tr

● Nerin Bahçeciler

Department of Child Health and Diseases, Division of Allergy
and Immunology, Near East University Faculty of Medicine,
Nicosia, Cyprus
nerin74@gmail.com

● Ömer Taşargöl

Department of Anesthesiology and Reanimation, Dr. Burhan
Nalbantoğlu State Hospital, Nicosia, Cyprus
omertasargol@yahoo.com

● Özlem Balcıoğlu

Department of Cardiovascular Surgery, Near East University
Faculty of Medicine, Nicosia, Cyprus
ozlem.balcioğlu@neu.edu.tr

● Sinem Şıgıt İkiz

Department of Radiology, Dr. Burhan Nalbantoğlu State
Hospital, Nicosia, Cyprus
sinemsigit@gmail.com

● Uğurcan Balyemez

Department of Radiology, Near East University Faculty of
Medicine, Nicosia, Cyprus
ubalyemez@gmail.com

● Umut Maraşuna

Department of Endocrinology, Dr. Burhan Nalbantoğlu State
Hospital, Nicosia, Cyprus
umutmousa@yahoo.co.uk

● Zeynep Taşargöl

Department of Obstetrics and Gynaecology, Dr. Burhan
Nalbantoğlu State Hospital, Nicosia, Cyprus
zeynepyt84@hotmail.com

CYPRUS

JOURNAL OF MEDICAL SCIENCES

Indexed in Web of Science

Volume: **11** | Issue: **5** | October 2026

EDITORIAL BOARD

National Advisory Board

Ali Ulvi Önder

Department of Urology, Near East University School of Medicine, Nicosia, Cyprus

Ayşe Gökyiğit

Department of Pharmaceutical Services of the Ministry of Health, Nicosia, Cyprus

Beste Kamiloğlu

Department of Orthodontics, Near East University School of Dentistry, Nicosia, Cyprus

Bülent Haydar

Private Clinic of Maxillofacial Surgery, Nicosia, Cyprus

Doğan Ceyhan

Department of Ophthalmology, Near East University School of Medicine, Nicosia, Cyprus

Düriye Deren Oygur

Department of Nephrology, Dr. Burhan Nalbantoğlu State Hospital, Nicosia, Cyprus

Ender Volkan

Cyprus International University School of Pharmacy, Nicosia, Cyprus

Erdem Beyoğlu

Barış Mental and Neurological Disorders State Hospital, Nicosia, Cyprus

Fatma Deniz

Department of Dermatology, Girne Akçiçek State Hospital, Girne, Cyprus

Filiz Besim

Private Clinic of Maxillofacial Surgery, Nicosia, Cyprus

Gamze Mocan Kuzey

Department of Pathology and Cytology, Near East University School of Medicine, Nicosia, Cyprus

Gönül Küçük

Department of Pediatric Surgery, Dr. Burhan Nalbantoğlu State Hospital, Nicosia, Cyprus

Gülşen Bozkurt

Private Clinic of Hematology, Nicosia, Cyprus

Hanife Erçal Ezgi

Department of Dermatology, Dr. Burhan Nalbantoğlu State Hospital, Nicosia, Cyprus

Hasan Besim

Department of General Surgery, Near East University School of Medicine, Nicosia, Cyprus

Hasan Mete İnancı

Private Clinic of Otorhinolaryngology, Nicosia, Cyprus

İdris Deniz

Department of Forensic Medicine, Dr. Burhan Nalbantoğlu State Hospital, Nicosia, Cyprus

İsmet Başar

Department of Urology, Dr. Burhan Nalbantoğlu State Hospital, Nicosia, Cyprus

Kaan Erler

Department of Orthopaedics, Near East University School of Medicine, Nicosia, Cyprus

Kenan Arifoğlu

Department of Plastic and Reconstructive Surgery, Dr. Burhan Nalbantoğlu State Hospital, Nicosia, Cyprus

Kerem Terali

Department of Medical Biochemistry, Near East University School of Medicine, Nicosia, Cyprus

Mehmet İnan

Department of General Surgery, Private Magusa Medicine Center, Famagusta, Cyprus

CYPRUS

JOURNAL OF MEDICAL SCIENCES

Indexed in Web of Science

Volume: 11 | Issue: 5 | October 2026

EDITORIAL BOARD

Meltem Nalça

Department of Radiation Oncology, Near East University School of Medicine, Nicosia, Cyprus

Murat Uncu

Department of Biochemistry, Near East University School of Medicine, Nicosia, Cyprus

Mustafa Kalfaoğlu

Department of General Surgery, Magusa State Hospital, Famagusta, North Cyprus

Mustafa Taşeli

Department of Ophthalmology, Near East University School of Medicine, Nicosia, Cyprus

Nahide Gökçora

Department of Nuclear Medicine, East Mediterranean University School of Medicine, Famagusta, Cyprus

Ozan Emiroğlu

Department of Cardiovascular Surgery, Dr. Burhan Nalbantoğlu State Hospital, Nicosia, Cyprus

Özay Önöral

Department of Prosthetic Medical Therapy, Near East University Faculty of Dentistry, Nicosia, Cyprus

Serap Soytaç İnancı

Private Clinic of Endocrinology and Metabolic Diseases and Internal Medicine, Nicosia, Cyprus

Sevda Lafcı

Department of Anatomy, Near East University School of Medicine, Nicosia, Cyprus

Sezgin Handan

Department of Nursing, Eastern Mediterranean University School of Health Sciences, Famagusta, Cyprus

Sibel Tozaki

Department of Dermatology, Dr. Burhan Nalbantoğlu State Hospital, Nicosia, Cyprus

Songül Acar Vaizoğlu

Department of Public Health, Near East University School of Medicine, Nicosia, Cyprus

Süha Akpınar

Department of Radiology, Near East University School of Medicine, Nicosia, Cyprus

Şanda Çalı

Department of Public Health, Near East University School of Medicine, Nicosia, Cyprus

Tarık İzbul

Department of General Surgery, Dr. Burhan Nalbantoğlu State Hospital, Nicosia, Cyprus

Tevfik Eker

Department of General Surgery, Private Magusa Medicine Center, Famagusta, Cyprus

Tijen Ataçağ

Department of Obstetrics and Gynecology, Near East University School of Medicine, Nicosia, Cyprus

Turgay Akalın

Private Clinic of Neurology, Nicosia, Cyprus

Ülvan Özad

Department of Plastic and Reconstructive Surgery, Near East University School of Medicine, Nicosia, Cyprus

CYPRUS

JOURNAL OF MEDICAL SCIENCES

Indexed in Web of Science

Volume: **11** | Issue: **5** | October 2026

EDITORIAL BOARD

International Advisory Board

A.C. Joao Lima

Department of Radiology, Johns Hopkins Medicine, Baltimore, USA

Aliye Özenoğlu

Department Nutrition and Dietetics, Üsküdar University School of Health Science, İstanbul, Türkiye

Alp Usubütün

Department of Pathology, Hacettepe University School of Medicine, Ankara, Türkiye

Alper Sertçelik

Department of Cardiology, Sanko University School of Medicine, Gaziantep, Türkiye

Ayla Ünsal

Department Of Nursing, Ahi Evran University School Of Health, Kırşehir, Türkiye

Ayşe Nihal Demircan

Department of Ophthalmology, Çukurova University School of Medicine, Adana, Türkiye

Aytekin Besim

Private Clinic of Radiology, Ankara, Türkiye

Bengi Semerci

Department of Psychiatrist, Institute of Bengi Semerci, İstanbul, Türkiye

Barış Doğu Yıldız

Department of General Surgery, Ankara Numune Research and Training Hospital, Ankara, Türkiye

Çağrı Büke

Department of Infectious Diseases and Clinical Microbiology, Yeditepe University School of Medicine, İstanbul, Türkiye

Cem Ertan

Department of Emergency Medicine, Akdeniz University School of Medicine, Antalya, Türkiye

Cem Terzi

Department of General Surgery, Dokuz Eylül University School of Medicine, İzmir, Türkiye

Coşkun Yorulmaz

Department of Forensic Medicine, İstanbul University Cerrahpaşa School of Medicine, İstanbul, Türkiye

Dilek Yavuz

Department of Internal Medicine and Endocrinology Section, İstanbul University School of Medicine, İstanbul, Türkiye

Ebru Yılmaz Yalçınkaya

Department of Physical Therapy and Rehabilitation, Gaziosmanpaşa Taksim Research and Training Hospital, İstanbul, Türkiye

Elif Arı Bakır

Department of Nephrology, Kartal Dr. Lütfi Kırdar Training Hospital, İstanbul, Türkiye

Egemen İdiman

Department of Neurology, Dokuz Eylül University School of Medicine, İzmir, Türkiye

Emre Canda

Department of General Surgery, Dokuz Eylül University School of Medicine, İzmir, Türkiye

Erkan Göksu

Department of Emergency Medicine, Akdeniz University School of Medicine, Antalya, Türkiye

Erol Baysal

Dubai Genetic and Thalassemia Center, Dubai Health Authority, Dubai, UAE

Fatih Köse

Department of Oncology, Başkent University School of Medicine, Adana Search and Practise Hospital, Adana, Türkiye

CYPRUS

JOURNAL OF MEDICAL SCIENCES

Indexed in Web of Science

Volume: 11 | Issue: 5 | October 2026

EDITORIAL BOARD

Fazıl Tuncay Aki

Department of Urology, Head of Transplantation Unit,
Hacettepe University School of Medicine, Ankara, Türkiye

Funda Tuğcu

Department of Orthodontics, Ankara University School of
Dentistry, Ankara, Türkiye

Gökhan Berkтуğ Bahadır

Department of Pediatric Surgery, Mersin University School of
Medicine, Mersin, Türkiye

Gülnur Göllü Bahadır

Department of Pediatric Surgery, Ankara University School of
Medicine, Ankara, Türkiye

Gökhan Nergizoğlu

Department of Internal Medicine-Nephrology, Ankara University
School of Medicine, Ankara, Türkiye

Gölge Acaroğlu

Private Clinic of Ophthalmology, Ankara, Türkiye

Hür Hassoy

Department of Public Health, Ege University School of Medicine,
İzmir, Türkiye

Hakan Altay

Department of Cardiology, Başkent University İstanbul Hospital,
İstanbul, Türkiye

Hüseyin Bakkaloğlu

Department of General Surgery, İstanbul University School of
Medicine, İstanbul, Türkiye

Hüseyin Mertsoylu

Department of Oncology, Başkent University School of Medicine,
Adana Search and Practise Hospital, Adana, Türkiye

İlhami Kuru

Department of Orthopedics and Traumatology, Başkent
University School of Medicine, Ankara, Türkiye

Kemal Bakır

Department of Pathology, Gaziantep University School of
Medicine, Gaziantep, Türkiye

Kemal Dolay

Department of General Surgery, Bezmialem Vakıf University,
Bezmialem Hospital, İstanbul, Türkiye

Kürşad Türksen

Samuel Lunenfeld Research Institute, Mount Sinai Hospital
University of Toronto, Toronto, Canada

Lale Tokgözoğlu

Department of Cardiology, Hacettepe University School of
Medicine, Ankara, Türkiye

Levent Sennaroğlu

Department of Otorhinolaryngology, Hacettepe University
School of Medicine, Ankara, Türkiye

Mazhar Tokgözoğlu

Department of Orthopaedics and Traumatology, Hacettepe
University School of Medicine, Ankara, Türkiye

Melih Atahan Güven

Department of Gynecology and Obstetrics, Acıbadem University
School of Medicine, İstanbul, Türkiye

Mustafa Camgöz

Department of Life Sciences, Imperial College School of Natural
Sciences, London, United Kingdom

Müfit Akyüz

Department of Physical Therapy and Rehabilitation, Karabük
University School of Medicine, Karabük, Türkiye

Müslime Akbaba

Department of Ophthalmology, Acıbadem University School of
Medicine, İstanbul, Türkiye

Mustafa Sertaç Yazıcı

Department of Urology, Hacettepe University School of
Medicine, Ankara, Türkiye

CYPRUS

JOURNAL OF MEDICAL SCIENCES

Indexed in Web of Science

Volume: **11** | Issue: **5** | October 2026

EDITORIAL BOARD

Neval Duman

Department of Internal Medicine-Nephrology, Ankara University
School of Medicine, Ankara, Türkiye

Nihat Yavuz

Department of General Surgery, İstanbul University School of
Medicine, İstanbul, Türkiye

Nilgün Kapucuoğlu

Department of Pathology, Acıbadem University School of
Medicine, İstanbul, Türkiye

Nilüfer Rahmioğlu

Department of Genetics, University of Oxford School of
Medicine, Oxford, United Kingdom

Nuray Başsüllü Kara

Department of Pathology, Acıbadem University School of
Medicine, İstanbul, Türkiye

Nuri Özgirgin

Department of Otorhinolaryngology, Bayındır Hospital, Ankara,
Türkiye

Orçun Şahin

Department of Orthopedics and Traumatology, Başkent
University School of Medicine, Ankara, Türkiye

Oytun Erbaş

Department of Experimental Medicine, The Scientific and
Technological Research Council (TUBITAK-Martek) of Türkiye, IL,
USA

Özgür Deren

Department of Obstetrics and Gynecology, Division of Maternal
Fetal Medicine, Hacettepe University, Ankara, Türkiye

Özgür Özyılkan

Department of Oncology, School of Medicine, Başkent University
Adana Search and Practise Hospital, Adana, Türkiye

Peyman Yalçın

Department of Physical Therapy and Rehabilitation, Ankara
University School of Medicine, Ankara, Türkiye

Pınar Zeyneloğlu

Department of Anesthesiology and Reanimation, Başkent
University, Ankara Hospital, Ankara, Türkiye

Ralph Tufano

Department of Otolaryngology-Head and Neck Surgery, Johns
Hopkins Medicine, Baltimore, USA

Rahmi Kılıç

Department of Otorhinolaryngology, Kırıkkale University School
of Medicine, Kırıkkale, Türkiye

Salih Marangoz

Department of Orthopaedics and Traumatology, Acıbadem
Mehmet Ali Aydınlar University School of Medicine, İstanbul,
Türkiye

Selçuk İnanlı

Department of Otorhinolaryngology, Head and Neck Surgery,
Marmara University School of Medicine, İstanbul, Türkiye

Serap Öztürkcan

Department of Dermatology, Celal Bayar University School of
Medicine, Manisa, Türkiye

Serkan Durdu

Department of Cardiovascular Surgery, Cebeci Kardiak Center,
Ankara University School of Medicine, Ankara, Türkiye

Serkan Sertel

Department of Otorhinolaryngology, University of Heidelberg
Neuenheimer Feld, Heidelberg, Germany

Serpil Altındoğan

Department of Oral Maxillofacial Surgery, Ankara University
School of Dentistry, Ankara, Türkiye

Server Serdaroğlu

Department of Dermatology, İstanbul University Cerrahpaşa
School of Medicine, İstanbul, Türkiye

Şaziye Şahin

Department of Anesthesiology and Reanimation, Gazi University
Dental School of Dentistry, Ankara, Türkiye

CYPRUS

JOURNAL OF MEDICAL SCIENCES

Indexed in Web of Science

Volume: 11 | Issue: 5 | October 2026

EDITORIAL BOARD

Teslime Atlı

Department of Geriatrics, Ankara University School of Medicine,
Ankara, Türkiye

Tolga Karcı

Department of Orthopaedics and Traumatology, İzmir Şifa
University İzmir, Türkiye

Ufuk Ateş

Department of Pediatric Surgery, Ankara University School of
Medicine, Ankara, Türkiye

Ufuk Erginoğlu

Department of Neurological Surgery, University of Wisconsin,
School of Medicine and Public Health, Madison, USA

Vedat Göral

Department of Gastroenterology, İstanbul Medipol University
School of Medicine, İstanbul, Türkiye

Vural Fidan

Department of Otorhinolaryngology, Yunus Emre State Hospital,
Eskişehir, Türkiye

Yeşim Sağlıcan

Department of Pathology, Acıbadem University School of
Medicine, İstanbul, Türkiye

Please refer to the journal's webpage (<https://cyprusjmedsci.com/>) for "Aims and Scope" "Journal Policies" and "Instructions to Authors". The editorial and publication process of the Cyprus Journal of Medical Sciences 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. Cyprus Journal of Medical Sciences is indexed in Clarivate Analytics – **ESCI**, **DOAJ**, **TR Dizin**, **EBSCO**, **GALE**, **Embase**, **CABI**, **J-Gate**, **Türkiye Atıf Dizini**, **Türk Medline**, **Idealonline**, **CNKI**, **Sudoc**, and **BASE**.

Owner: Cyprus Turkish Medical Association

Responsible Manager: Sonuç Büyük



Publisher Contact

Address: Molla Gürani Mah. Kaçamak Sk. No: 21/1 34093
İstanbul, Türkiye

E-mail: info@galenos.com.tr/yayin@galenos.com.tr

Web: www.galenos.com.tr Publisher Certificate Number: 14521

Publication Date: October 2026

E-ISSN: 2536-507X

ISSN: 2149-7893

International scientific journal published bi-annually.

CYPRUS

JOURNAL OF MEDICAL SCIENCES

Indexed in Web of Science

Volume: **11** | Issue: **5** | October 2026

CONTENTS

REVIEWS

334 Tuberculosis and Migration: From Global to Local Perspective

Ayşe Baha, Fatma Canbay; Nicosia, Cyprus

RESEARCH ARTICLES

342 Evaluation of Depressive Symptoms in Children Aged 3-6 Years: A Parent-Report Study

Mine Cengiz, Yusuf Aktepe, Burcu Demir Gökmen; Erzurum, Ağrı, Türkiye

353 The Impact of Laparoscopic and Transvaginal Vaginal Cuff Closure Techniques on Surgical and Functional Outcomes in Total Laparoscopic Hysterectomy

Mehmet Emre Peker, Ufuk Atlıhan, Can Ata, Duygu Uçar Kartal, Haydar Kaya; Manisa, İzmir, Türkiye

359 Polycystic Ovary Syndrome Detection Using Gradient Boosting Algorithms and Explainable Artificial Intelligence: A Comparative Study

Ümran Karabulut Doğan, Şeyma Yaşar, Erhan Hüseyin Cömert, Erol Karakaş, Ozan Doğan; Kayseri, Malatya, İstanbul, Türkiye

369 Evaluating Large Language Models' Responses to Common Queries on Retinal Surgery: Readability and Accuracy Analysis

Erkan Duman, Aydın Maçın, Ayşe Ergin; Erbil, Iraq; Diyarbakır, İstanbul, Türkiye

378 Asymptomatic Anterior Tarsal Tunnel Syndrome: An Underdiagnosed Entity

Filiz Azman-İste, Berfe Simya Özbek; İstanbul, Türkiye

384 Temporomandibular Joint Arthrocentesis: A Bibliometric Analysis From 1989 to 2024

Mustafa İşleyen, Asenur Aydemir; Burdur, Türkiye

393 Analysis of the Relationship Between Disease Self-Management and Psychosocial Adjustment of Celiac Patients

İlknur Bodur, Hilal Tüzer; Ankara, Türkiye

401 Compliance with Tobacco Control Law in Hospitals of Nicosia: An Observational Study

Özen Aşut, Songül Acar Vaizoğlu, Şanda Çalı, Ahmed Elamin Ahmed Elamin, Mohamed Ebied, Enver Abdullaiev, Karam Chaddad, Muhnd Muhammad Algawi, Shania Sheka, Abdullatif Adnan İbrahim, Tuqa Mohammad Aldowery, Gülşah Bahadır, Mehmet Utku Durum, Alper Çeliker; Nicosia, Cyprus

409 Associations Between Conventional Breast MRI Signal Characteristics and Immunohistochemical Biomarkers in Breast Cancer: A Retrospective Exploratory Study

Uğurcan Balyemez, Sinan Akay, Hatice Tuba Sanal, Mustafa Taşar; Nicosia, Cyprus; Iowa, United States; İstanbul, Türkiye

417 Screen Exposure and Socio-Demographic Characteristics of Children Aged 2-6 Years Diagnosed with Speech and Language Disorder: A Descriptive Study

Bari Ay, Umut Balatacı, Muzaffer Tunahan Gedikçi; Karabük, İstanbul, Ankara, Türkiye

424 The Epidemiological and Clinical Profile of the Patients Having Oocyte Donation Treatment in Northern Cyprus

Burcu Özbakır; Nicosia, Cyprus

430 Global Trends and Emerging Research Themes in Patient-Reported and Patient-Centered Outcomes after Cataract Surgery: A Bibliometric and Science Mapping Analysis

Dilber Keskinel, Gülcan Dürüst Sakallı; Nicosia, Famagusta, Cyprus

Tuberculosis and Migration: From Global to Local Perspective

✉ Ayşe Baha¹, ✉ Fatma Canbay²

¹Department of Pulmonology, Near East University Faculty of Medicine, Nicosia, Cyprus

²Clinic of Chest Diseases, Dr. Burhan Nalbantoğlu State Hospital, Nicosia, Cyprus

Abstract

Tuberculosis (TB) continues to be a significant global public health issue. It is estimated that around 25% of the world's population is infected with *Mycobacterium tuberculosis*, and TB remains one of the top causes of death worldwide. The World Health Organization's (WHO) latest report indicates that in 2023, there were 10.8 million TB cases and 1.25 million TB-related deaths globally. That year, 8.2 million new TB cases were documented, representing an increase compared with previous years. Population movement is recognized as a significant factor affecting the epidemiology of TB. Migration has increasingly posed a global challenge in social, economic, and public health domains. By the close of 2024, around 118.6 million individuals worldwide had been forcibly displaced due to conflict, violence, persecution, or violations of human rights. Furthermore, the number of people internally displaced by conflict or violence rose by 10% from 2022 to 2023. Ongoing conflicts, economic inequality, and educational opportunities are anticipated to continue to drive migration in the years ahead. WHO underscores the necessity of global cooperation in the screening, diagnosis, treatment, and monitoring of TB among migrants and refugees. Recently, our country has experienced an increase in migration driven by job and educational opportunities. While TB cases have decreased among the native population, there has been an uptick in cases among those born abroad. In light of global patterns and WHO guidelines, it is crucial to reevaluate and enhance our country's TB screening strategies for immigrants within a robust public health framework.

Keywords: Migrant, migration, refugees, tuberculosis

INTRODUCTION

Tuberculosis (TB) is an infectious, able to be prevented and treated, disease caused by the *Mycobacterium tuberculosis* bacillus.¹ Although it most commonly influences the lungs [pulmonary TB (PTB)], it may also affects many organs extrapulmonary TB (EPTB).² TB, which causes morbidity and mortality in millions of people all over the world every year, is the tenth leading reason of death.³

Most people who have TB disease live in 30 countries with a high TB burden. Most TB patients live in 30 countries, that account for 87 percent of the world's population in 2023, with a high TB burden.³ The countries that highest incidence of TB are Pakistan (6.3%), Philippines (6.8%), China (6.8%), Indonesia (10%) and India (26%), and these countries account for 56 percent of the global population (Figure 1).³

According to the World Health Organization (WHO), there were 10.8 million TB patients globally in 2023 and 1.25 million deaths happened that year.³ The number of new TB cases in 2023 was 8.2 million people.³ This number was 7.1 million in 2019, 5.8 million in 2020, 6.4 million in 2021, and 7.5 million in 2022.³ Possible reasons for the increase in 2022 and 2023 include global migration and difficulties accessing healthcare services during the coronavirus disease pandemic.

TB bacilli are transmitted from an infectious person (TB patient) to a healthy person via the airways.⁴ The most infectious patients are those with positive sputum microscopy for acid-resistant bacilli and cavitated pulmonary and laryngeal TB. Smear negative TB patients are less infectious.⁴ The risk of transmission is high for people with close, long-term contact with a patient with TB. These may be family members,

To cite this article: Baha A, Canbay F. Tuberculosis and migration: from global to local perspective. Cyprus J Med Sci. 2026;11(5):334-341

ORCID IDs of the authors: A.B. 0000-0002-9649-250X; F.C. 0000-0001-8127-9995.



Corresponding author: Ayşe Baha
E-mail: dr_ayse demir@hotmail.com
ORCID ID: orcid.org/0000-0002-9649-250X

Received: 17.07.2025
Accepted: 02.08.2026
Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.
This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

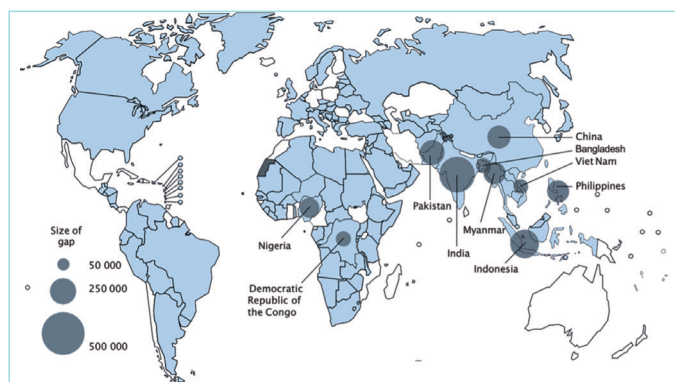


Figure 1. The ten countries with the largest gaps between case notifications of people newly diagnosed with TB and the best estimates of TB incidence, 2023.³

TB: Tuberculosis.

housemates, or work colleagues. Other risk factors for transmission include poorly ventilated environments, crowded environments, not recognising the source case, not being treated or receiving incomplete treatment, increased cough severity of the source case and drug-resistant variant.⁴

It is estimated that one quarter of the world's population is infected with TB today.³ Being infected with TB bacilli does not mean being a TB patient, it only means that the body has been exposed to the bacilli.⁵ This is called latent TB infection (LTBI). Over time, latent infection can progress to TB disease, which is the principal clinical problem. The risk of individuals with LTBI developing TB is highest in the first 2 years (roughly 5%) and this risk decreases over time.^{3,5} These individuals have a lifetime risk of developing TB of 5 to 15 percent.⁵

TB is a treatable disease, but treatment may be inadequate due to factors such as prolonged therapy, the need for multiple drugs, and difficulty accessing medication. Drug-resistant TB (DR-TB) may develop over time in people who fail to complete their treatment, and these patients constitute the most costly and deadly group in medical and socioeconomic terms. Strategies and targets to combat TB are reviewed annually by WHO and other international health organisations, and substantial efforts are made in this regard. Although it will take a long time to reach the targets of the end TB (END-TB) strategy conducted by WHO, the 2025 results indicate that TB incidence has been reduced by 50 percent and deaths from TB have been reduced by 75 percent over the last 10 years. However, there are serious challenges to achieving the objective of TB eradication, especially in countries with high TB incidence. The fundamental problem is the cost required for diagnosis and treatment.³

History of the Spread of Tuberculosis

The existence of TB, which is thought to have been present on Earth since the age of the dinosaurs, has also been proven in prehistoric Africa.⁶ The disease of TB in humans started with the domestication of cattle.⁷ Humans who began living with cattle encountered *Mycobacterium bovis*, the "bovine TB bacillus", for the first time through consumption of animal milk and meat and became ill. Over time, this bacillus has acquired genetic changes consistent with the more virulent human strain, *Mycobacterium tuberculosis* *humanus*. As people started to live in communities of more than 25,000 people, the spread of TB became

easier.⁸ Through the intercontinental migrations of humans, the TB bacillus travelled to America through the Bering Strait, which was opened 9000 years ago.⁸ Today, migration remains a major risk factor for the spread of TB worldwide.

What is Migration, Migrant and Refugee?

The movement of a person or people from a country, locality, place of residence, etc., to settle in another is called "migration".⁹ Since the earliest periods of history, people have migrated. Today, factors such as new job opportunities, education, wars, human rights violations, and terrorism cause migration. Currently, 281 million people (3.6%) worldwide are migrants.³ In addition to its economic, social, and psychological effects, migration facilitates the spread of infectious diseases such as TB. This further increases its significance.

Refugees are specified as persons who are outside their country and unable to return because of a well-founded fear of persecution on account of their religion, nationality, race or political opinion.¹⁰ However, migrants are defined as individuals who move to another country to improve their academic, financial or social situations and enhance prospects for themselves.¹¹ TB is an important reason of morbidity and mortality amongst refugees and migrants.¹² By the end of 2024, 118.6 million people globally were forcefully displaced because of conflict, persecution or human rights violations.¹³ In 2023, more than half of all refugees were estimated to have come from just three countries: Ukraine, Afghanistan, and Syria (respectively 5.9, 6.1, and 6.7 million).¹⁴ The total number of refugees all over the world has increased by at least three times in the last decade, with three-quarters living in low- and middle-income countries.¹⁵ Additionally, the number of internally displaced persons as a result of violence or conflict has increased by 10 percent from 2022 to 2023.¹⁵

Overview of Tuberculosis Risk in Migrants and Refugees

Refugees and migrants are the people most at risk of contracting infectious diseases such as TB due to their living conditions.¹⁶ Migrants and refugees are at greater risk of contracting TB than the native population. This is associated with factors such as TB prevalence in their country, migration route, frequency of visiting their home country, living conditions and duration of residence in the host country.^{17,18} The impact of migration on the epidemiology of TB is significant.¹⁹ The arrival of large groups of refugees significantly increases the burden and cost of TB. This also affects TB control efforts in host countries that accept refugees.¹⁸ Because of most migrants come from low-income countries with a high disease burden, this condition reveals serious challenges for global control and elimination of TB.²⁰ In migrants, breakdowns and blocking in access to health services is main risk in terms of spread of TB.³

Tuberculosis Epidemiology in Migrants and Refugees Worldwide

A systematic review published in 2022 summarised the eminence of TB among migrants and refugees.²¹ This review analysed European countries and found that migration from Syria to other countries in 2011 increased TB incidence in both border countries and elsewhere in Europe. Similarly, there were large migrations from African and Asian countries to Europe in those years, and it was stated that these also contributed to the incidence of TB.²¹ Türkiye borders Syria, and it was the largest recipient of migrants during the 2011 Syrian crisis. The rate of imported TB cases in Türkiye increased from 1.3% to 6.8% from 2011

to 2015, reflecting a important increase.²² In Hatay, which borders Syria, the prevalence of TB patients increased from 25.1% to 34.6% between 2010 and 2013.²²

The total number of Syrians in Lebanon is estimated to be higher than 1.5 million, and it represents more than a quarter of the country's population. Unlike other countries, Lebanon has no official refugee camps. That's why refugees are randomly distributed throughout the country.²³ TB incidence in Lebanon declined until 2011. Yet, since the beginning of the Syrian crisis, there has been remarkable rising in TB patients in Lebanon.^{22,24} Before the Syrian crisis, the prevalence of TB cases in Lebanon was 33 percent, but gradually has risen to 53 percent from 2010 to 2015.²²

Another meta-analysis on TB among migrants²¹ indicated that TB patients were aged between 13 and 54 years and that the disease was more prevalent among men than women. Most refugees and migrants with TB are from Asian and African countries. The most frequently reported Asian countries were Vietnam, Thailand, the Philippines, India, Pakistan, and Syria, while the most frequently reported African countries were Somalia and Eritrea.²¹ In the meta-analysis, the incidence and prevalence of TB in migrants and refugees were found to vary across countries. Canada has the lowest TB rate among refugees at 19 per 100,000,²⁵ while the Netherlands has the highest at 754 per 100,000.²⁶ Italy has the highest prevalence in Europe. The prevalence in immigrants is 80 times higher than the estimated prevalence rate in Italy (6.7 per 100,000 population).²¹

In 2020, more than 9,000 foreign-born people in the European Union/European Economic Area (EU/EEA) developed TB.²⁷ The number of Ukrainians developing TB in the EU/EEA rose from 164 in 2020 to 780 in 2022. While 27 of the cases in 2020 were multidrug/rifampin-resistant-TB, this number increased to 194 in 2022.²⁸ This increase indicates significant changes in the epidemiology of TB in the EU/EEA due to migration in recent years.

The most common type of TB among migrants and refugees is PTB. In fact, this group has the highest risk of developing EPTB.²⁹ EPTB is important because diagnosis/treatment delays are significantly greater compared to PTB.²⁹

According to the most recent large meta-analysis, diagnostic delay for TB is greater among migrants than non-migrants. However, the time to initiation of treatment is shorter among migrants than among non-migrants.³⁰⁻³² The main reasons for delayed diagnosis among migrants with TB are distance to health centres, the cost of healthcare services and concerns about stigmatisation.³³ Recent study including only immigrants analysed delays of diagnosis and treatment, and it founds that delays varied according to country of origin and level of education.²⁹ In a meta-analysis including 60,802 migrants with TB and 221,002 non-migrants with TB, negative treatment outcomes such as treatment failure, treatment refusal and loss to follow-up were higher in migrants [odds ratio: 1.76 (95% confidence interval: 1.01-3.03)].³³

Drug-Resistant Tuberculosis and Its Role in Migration

DR-TB refers to a disease that is resistant to one or more drugs used in treatment. It is analysed in 5 categories by WHO; isoniazid resistant TB, RR-TB, multidrug resistant-TB (MDR-TB), rifampicin and fluoroquinolone resistant TB and rifampicin, fluoroquinolone and bedaquiline or linezolid resistant TB (XDR-TB).⁴

The most significant problems with DR-TB are treatment difficulty, increased risk of transmission, and high mortality. While the duration of treatment in classical PTB is 6 months with oral drugs, treatment for DR-TB takes years, and many drugs must be administered intravenously concurrently, thus requiring hospitalisation for months or even years. However, since this is not possible in many countries, these patients remain infectious without receiving treatment. Access to drug resistance testing is unavailable in many settings. Therefore, it is thought that most cases are undetected. Each undetected case continues to transmit the disease and, unfortunately, often results in death. The diagnosis and treatment of DR-TB depends on a lot of factors such as the success of the health system, integrated efforts of governments and international cooperation.

Access to drug-resistance testing is limited. Despite this, the prevalence of DR-TB is currently estimated at 5.5 percent and is on the rise (Figure 2).³ According to the WHO, only 25% of all DR-TB patients worldwide started treatment in 2023.³ 75% do not have access to testing and treatment. The population that accounts for this large difference lives in Ukraine, Myanmar, Pakistan, China, Indonesia, Philippines, India, Nigeria, Vietnam and South Africa.³ Today, these countries account for the highest number of immigrants to Europe and neighbouring countries. In a meta-analysis conducted in Europe, DR-TB was found to be 3.6 times more common in immigrants than in non-immigrants.³³

In China, Bangladesh and Russia, the rates of multidrug resistance in previously treated TB cases are 26%, 29%, 99%, respectively.³⁴ In Eastern Europe, the rate of multidrug resistance in previously treated TB cases is 69.0% in Belarus, 62.0% in Moldova and Estonia, 56.0% in Ukraine, 49.0% in Lithuania, 30% in Latvia and 23.0% in Bulgaria.^{35,36} Migrants from countries with high rates of DR-TB are increasingly settling in European countries with low TB incidence.

It has been reported that nearly three-quarters of MDR-TB cases detected in EU and EEA member states were identified in migrants.³⁴ In Germany, migrants account for 94% of MDR-TB cases, but only 58.7% of TB cases.³⁴ Worldwide, only half of MDR-TB patients successfully complete treatment, while a quarter disappear without any information on the treatment result.³ Only 26 percent of XDR-TB patients successfully complete treatment, while a quarter drop out of follow-up.³ This difference is greater in countries where drug resistance has a higher prevalence.³ The main factors that compromise adherence to MDR-TB treatment can be listed as multi-drug therapy, long-term treatment, and concerns about social stigma.³⁷⁻³⁹

Some data suggest that MDR-TB treatment success is lower among migrants than among the local population in low-incidence European countries. It is thought that this situation may be related to barriers such as language, health literacy and lack of access to healthcare services.³⁴

Human Immunodeficiency Virus-Tuberculosis Co-Infection and Its Role in Migration

The most common opportunistic disease in individuals with human immunodeficiency virus (HIV) is active TB and HIV is the strongest known risk factor for TB development.⁴⁰ The risk of developing active TB is twenty times higher in HIV-positive individuals with latent TB,⁴⁰ leading to a significant increase in morbidity and mortality in these cases.⁴⁰ Another reason why TB is important in people with HIV is that the drugs used in the treatment of both diseases interact with each other.⁴ The most significant risks for HIV-TB co-infection among

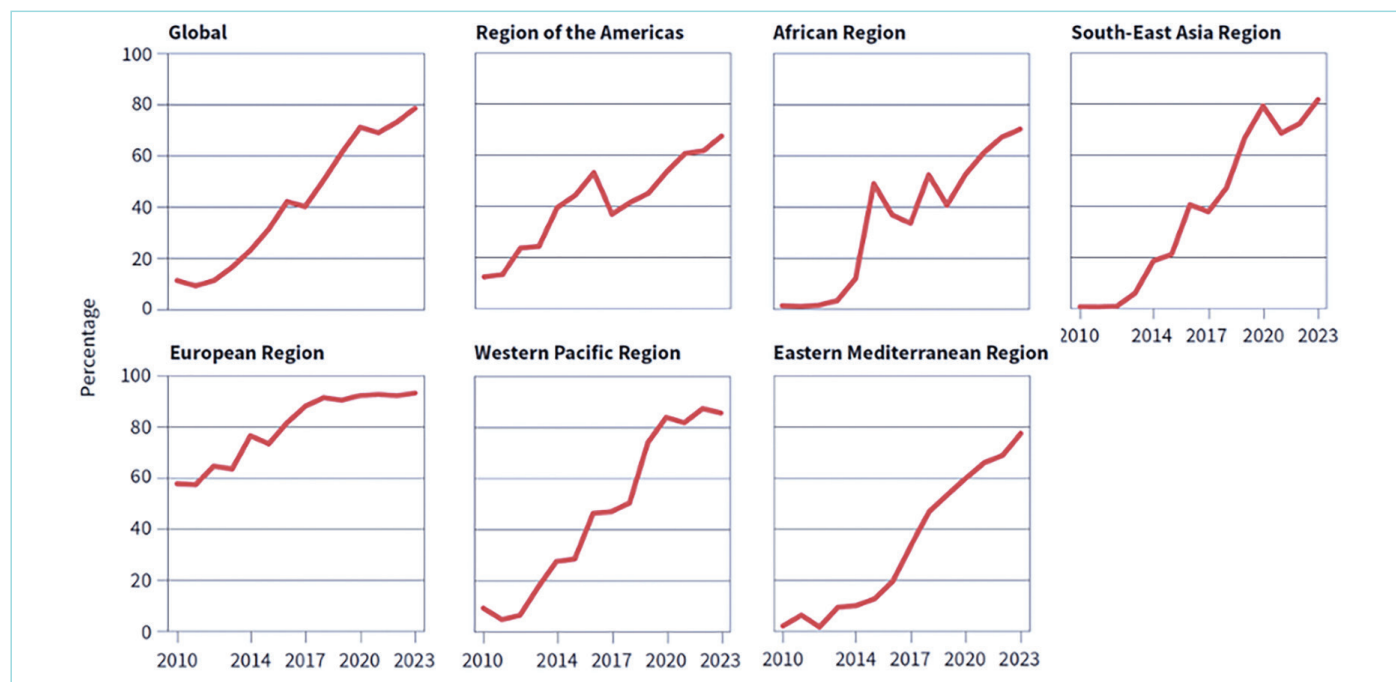


Figure 2. Percentage of people with RR-TB, globally and for WHO regions, 2010-2023.³

RR-TB: Rifampicin-resistant tuberculosis, WHO: World Health Organization.

migrants are poor socio-economic status, poor living conditions and limited access to healthcare services.⁴⁰ Despite this, studies on HIV-TB epidemiology in migrants are insufficient.

The development of antiretroviral treatments over the last few decades has reduced the incidence of HIV-TB co-infection, but the high mortality rate among these individuals remains a significant problem. HIV-TB constitutes 6.3% of all TB cases in 2022.³ Nearly 70% (460,000 cases in total) of HIV-TB cases in 2023 were detected in the African region, especially in Eastern and Southern Africa. In Thailand, it is estimated that unregistered migrants are 3 times more than registered migrants and 30% of TB cases are reported to be HIV-TB.³ Among people with HIV-TB in Portugal, 36.4% are immigrants and 63.6% are natives. Most immigrants (80.3%) came from African countries.⁴⁰

The prevalence of HIV-TB co-infection in Europe increased from 2008 to 2014 [3 in 7 studies from United Kingdom (UK), Spain, France and Portugal, HIV-TB co-infection was higher in immigrants (2.8%) than natives (2.3%)].⁴⁰ In 1994, the prevalence of HIV-TB co-infection among immigrants in Barcelona was 6.5%, rising to 37.1% in 2004, while in the UK it rose from 45% in 2001 to 78% in 2002.⁴⁰ The majority of migrants with HIV-TB in Europe originate from Africa (1-76.6%), Western Europe (3.2%), Eastern Europe (42.8%), South Asia/East Asia/Pacific (36.4%) and Latin America (2-31.7%).⁴⁰

The mortality rate of HIV-TB cases among immigrants has been reported inconsistently. Between 2001 and 2011, it was found that mortality rates among co-infected patients arriving in Germany from Sub-Saharan Africa were higher than those among natives.³ This situation is similar in Spain and Italy.⁴⁰

It is considered that there were 740,000 (95% UI: 550,000-950,000) cases with HIV-TB deaths in 2006 and this number decreased by nearly 78% to 160,000 (95% UI: 130,000-190,000) in 2023.³ The development

of antiretroviral therapies is the main reason for these reductions in incidence and mortality rates.

Tuberculosis Epidemiology in Turkish Republic of Northern Cyprus and the Impact of Migrants

In recent years, there has been a decrease in the number of TB patients in our country. According to official data, the number of cases decreased from 70 in 1989 to 32 in 2013, 35 in 2014, 39 in 2015, 20 in 2016, 29 in 2017 and 22 in 2018.⁴¹ Although the total number of TB cases has decreased, the vast majority occur among migrants. Eleven of the 22 cases reported in 2018 were migrants. The majority of these individuals were from the former Soviet Union, Asia, and Africa. In 2023, a total of 61 people were newly diagnosed with TB in the Turkish Republic of Northern Cyprus (TRNC). While only 3 of these patients were TRNC citizens, the remaining 58 patients were primarily from countries of the former Soviet Union, Asia, Africa, and the Republic of Türkiye. Twenty of the cases were female, and 41 were male. The majority of migrant TB cases (54, 88.5%) were PTB. In addition, bone TB was diagnosed in 3 cases, lymph node involvement in 1 case, and ocular TB in 1 case. These data indicate that the epidemiology of TB in the country is increasingly associated with migrant populations, and that extrapulmonary forms are included in the diagnostic spectrum.

In 2024, a total of 45 TB cases were recorded, of which only 2 were TRNC citizens. According to the available data for 2025, only 1 of the 2 cases detected so far has TRNC citizenship. These findings indicate that the majority of TB cases diagnosed in recent years occurred among migrants, with a marked downward trend in the number of local cases. This demonstrates the impact of migration on the epidemiology of TB. In addition, it highlights the need for a screening and control strategy that accounts for global migration.

Migration-Related Tuberculosis Screening Strategies

Current practices in the United States and Europe

Regular migrants from regions where TB is endemic have a similar risk of contracting TB as in their home countries, but it is generally higher than in the country where they will settle.⁴² For this reason, countries with low incidence rates require pre-migration screening, particularly for migrants arriving from high-incidence regions.⁴² The aim is to identify TB patients and migrants who are at high-risk of developing TB after migration.⁴²

Although TB screening strategies in migrants vary between countries and regions, WHO recommends systematic screening of migrants for active disease by chest x-ray (CXR) and tuberculin skin test (TST) or interferon- γ release test (IGRA) for latent TB infection (LTBI) in migrants from countries with high TB burden.^{3,42} The European Centre for Disease Prevention and Control (ECDC) recommends screening all migrants from high-incidence countries using CXR, TST or IGRA.^{3,42} The ERS report states that all migrants with a TB prevalence of more than 100 per 100,000 in the country of origin should be screened for active TB and LTBI on entry into the EU/EEA and the UK.³³

If an immigrant at risk of TB has not been screened for TB, it is recommended that TB or latent infection screening should be performed within the first 2 years after entering the country.³³ It is even advocated that migrants who are excluded from active TB should be given short-term preventive treatment regardless of LTBI status.³³ TB- screening strategies for migrants in various countries are summarised in Table 1.

Current practices in Turkish Republic of Northern Cyprus

In the TRNC, the TB Dispensary has been conducting regular TST for certain groups, especially primary-school students in grades 1 and 5, since 1976. Additionally, individuals are screened with the TST during military recruitment. Until 2023, migrants in the country who held a work permit were also screened for TB using the TST. However, since the second half of 2023, a screening method based on C-reactive protein (CRP) level has been adopted instead of the TST to screen migrants. Under this new practice, individuals with CRP levels above the normal reference range are further evaluated by CXR for PTB. In our country, people diagnosed with TB, especially those with highly infectious PTB,

are hospitalised and monitored to ensure isolation and treatment effectiveness. Following confirmation of the diagnosis, molecular genetic tests to detect rifampicin resistance are ordered for all patients. In our country, TB treatment is provided free of charge in line with public health policies, and the medicines are supplied by the Ministry of Health.

World Health Organization strategies in migration-associated tuberculosis

In 2014, the WHO launched the END-TB strategy. Its goal is to end the global TB epidemic by 2030 (Figure 3). At the 2023 United Nations meeting, a political declaration was issued, and heads of state made the following commitments:

- To provide TB prevention and treatment to 90% of those in need.
- To ensure that all TB patients undergo the WHO-recommended rapid diagnostic test.
- To ensure that all TB patients are afforded with a health and social support package.
- To ensure the development and use of at least one new, safe and effective TB vaccine.
- To provide the necessary funding for TB research until 2027.

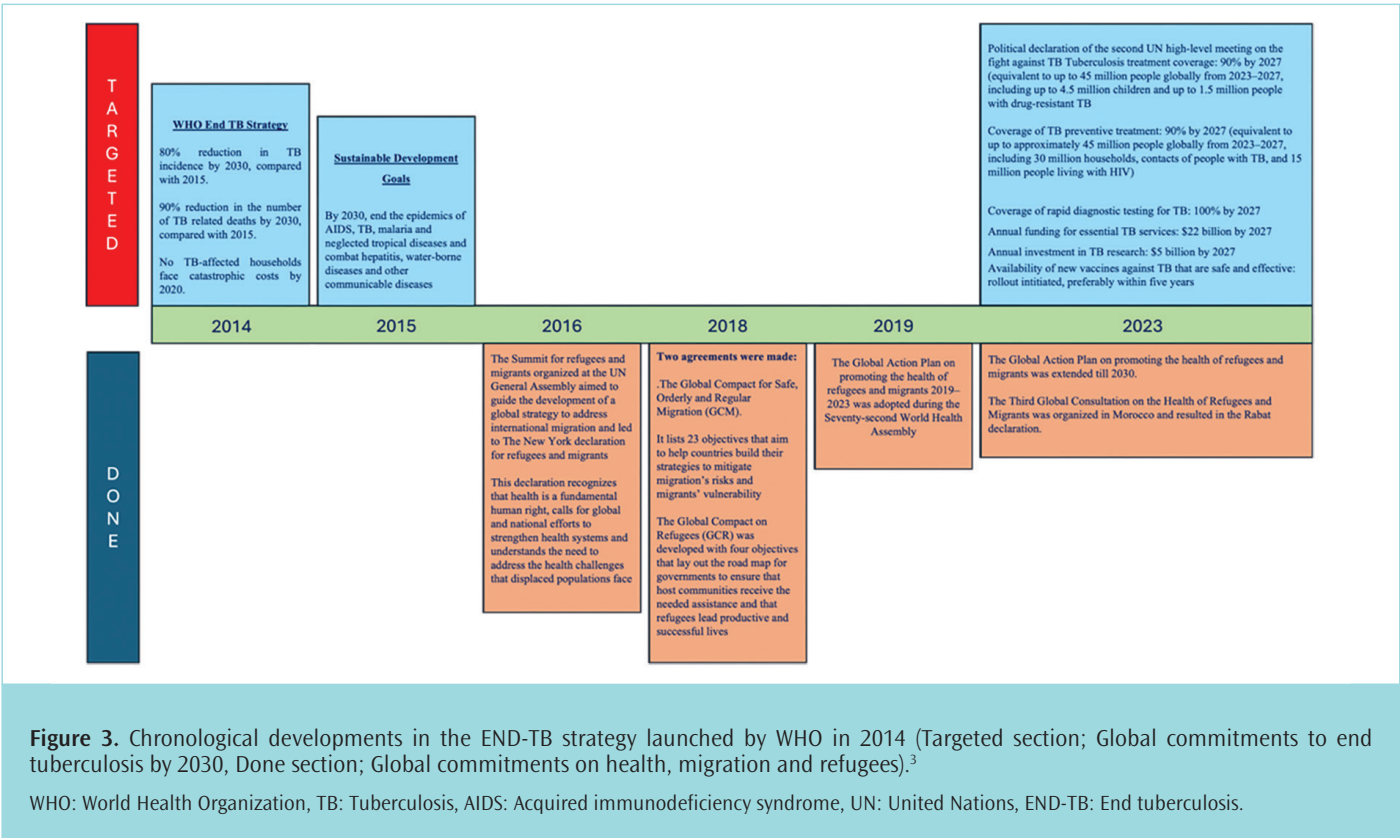
Implementation requirements in the Turkish Republic of Northern Cyprus

The PPD test used in TB screening in the TRNC is both a diagnostic tool with proven efficacy and a tool recommended by the WHO.³ However, as of 2023, screening practices based on CRP levels have replaced the PPD test. Since CRP is not a specific marker for TB and reflects general inflammatory responses, this approach poses a risk of both false positivity and delay in diagnosis.⁴³ This practice poses significant threats to public health and may lead to a weakening of epidemiological control. Experts consider the return to the TST-based screening method with high scientific accuracy, the strengthening of the infrastructure for the diagnosis-treatment process, and the establishment of a holistic national TB control programme as public health priorities. In this framework, migrants applying for work or residence permits, military

Table 1. Countries' TB screening strategies for migrants

Country	Population	Timing	Participation	Method
France	Children from countries with TB incidence >40 per 100 000 up to the age of 18 years Adults aged 18-40 years who live with children and come from countries with an incidence >100 per 100 000	On arrival or shortly after arrival	Voluntary	IGRA/TST
Italy	From countries with TB incidence >100 per 100,000	On arrival or shortly after arrival	Voluntary	IGRA/TST
Norway	From countries with TB incidence >200 per 100,000, Afghanistan and Eritrea, under the age of 35 years	On arrival or shortly after arrival	Mandatory	IGRA
Sweden	From countries with TB incidence >100 per 100,000 or with an exposure risk factor	On arrival or shortly after arrival	Voluntary	IGRA/TST
UK	Migrants aged 16-35 years who have been in the UK for up to 5 years and have lived for >6 months in high-incidence countries (≥ 150 per 100,000) or sub-Saharan Africa	Pre-entry	Voluntary	IGRA
Canada	All migrants	Pre-entry	Mandatory	Chest X-ray/IGRA or TST
US	Migrants aged ≥ 2 years, from countries with TB incidence ≥ 20 per 100,000	Pre-entry	Mandatory	Chest X-ray/IGRA
TRNC	Migrants aged >15 years, from countries with TB incidence <20 per 100,000	Pre-entry	Mandatory	Chest X-ray

TB: Tuberculosis, IGRA: Interferon-gamma release assay, TST: Tuberculin skin test, UK: United Kingdom, US: United States, TRNC: Turkish Republic of Northern Cyprus.



personnel, and student groups should be screened using TST for early detection of TB or LTBI; if the test is positive or there is clinical suspicion, they should be further evaluated by chest radiography for diagnostic confirmation.

CONCLUSION

This review has revealed significant differences between migrants and non-migrants in terms of TB prevalence, treatment and follow-up processes. Migrants seek treatment later, have a higher incidence of EPTB, and experience poorer treatment outcomes. The prevalence of MDR/RR-TB and TB/HIV is higher among migrants; however, it varies between countries. The general approach in countries with low TB incidence is to screen migrants for TB/LTBI, to provide a minimum TB care package for drug-sensitive and MDR/RR-TB, and to develop LTBI preventive treatment strategies for migrants with HIV co-infection. To improve TBTB outcomes among migrants, it is necessary to develop and implement care strategies tailored to this population, enhance the cultural competence of healthcare providers, and establish advanced support systems.

MAIN POINTS

- Global migration is reshaping tuberculosis (TB) epidemiology.
- Barriers to healthcare contribute to delayed TB diagnosis in migrant populations.
- Evidence-based TB screening improves early detection in high-risk populations.

- Current TB screening practices in the Turkish Republic of Northern Cyprus warrant an evidence-based revision.
- Equitable, non-stigmatising TB policies are essential for effective TB control.

Footnotes

Authorship Contributions

Concept: A.B., Design: A.B., F.C., Literature Search: A.B., F.C., Writing: A.B., F.C.

DISCLOSURES

Conflict of Interest: One author of this article, Ayşe Baha, is a member of the Editorial Board of the Cyprus Journal of Medical Sciences. However, she did not involved in any stage of the editorial decision of the manuscript. The editors who evaluated this manuscript are from different institutions. The other author declared no conflict of interest.

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

Declaration on the Use of Artificial Intelligence (AI): No artificial intelligence tools were used in the preparation of this manuscript.

REFERENCES

1. Meaza A, Tola HH, Eshetu K, Mindaye T, Medhin G, Gumi B. Tuberculosis among refugees and migrant populations: systematic review. PLoS One. 2022; 17(6): e0268696.

2. Centers for Disease Control and Prevention. Tuberculosis: basic TB facts. Atlanta, GA: CDC. Available from: <https://www.cdc.gov/tb/index.html>
3. World Health Organization. Global tuberculosis report 2024. Geneva: World Health Organization; 2024.
4. T.C. Sağlık Bakanlığı, Halk Sağlığı Genel Müdürlüğü. Tüberküloz Tanı ve Tedavi Rehberi. 3. baskı. Ankara: T.C. Sağlık Bakanlığı; 2026.
5. Getahun H, Matteelli A, Chaisson RE, Ravigione M. Latent mycobacterium tuberculosis infection. *N Engl J Med*. 2015; 372(22): 2127-35.
6. Hayman J. *Mycobacterium ulcerans*: an infection from Jurassic time? *Lancet*. 1984; 2(8410): 1015-6.
7. Keskinbora K. Another enemy in war: tuberculosis. *Lokman Hekim Journal*. 2016; 6(3): 174-84.
8. Daniel TM. The origins and precolonial epidemiology of tuberculosis in the Americas: can we figure them out? *Int J Tuberc Lung Dis*. 2000; 4(5): 395-400.
9. McAuliffe M, Oucho LA, editors. World Migration Report 2024. Geneva: International Organization for Migration (IOM); 2024.
10. Jackson IC. The 1951 Convention relating to the Status of Refugees: a universal basis for protection. *Int J Refugee Law*. 1991; 3(3): 403-13.
11. Urquia ML, Gagnon AJ. Glossary: migration and health. *J Epidemiol Community Health*. 2011; 65(5): 467-72.
12. Connolly, Máire A, Gayer, Michelle, Ottmani, Salah-Eddine, WHO Disease Control in Humanitarian Emergencies Programme & UNHCR (2007). Tuberculosis care and control in refugee and displaced populations : an interagency field manual / edited by M.A. Connolly, M. Gayer and S. Ottmani, 2nd ed. World Health Organization. Available from: <https://iris.who.int/handle/10665/43661>
13. World Health Organization, World Innovation Summit for Health. Innovative solutions for the elimination of tuberculosis among refugees and migrants: report of the 7th edition of the World Innovation Summit for Health. Geneva: World Health Organization; 2024.
14. UNHCR refugee data finder 2023. Available from: <https://www.unhcr.org/sites/default/files/2023-10/Mid-year-trends-2023.pdf>
15. United Nations High Commissioner for Refugees. Global Trends: Forced displacement in 2023. Copenhagen: United Nations High Commissioner for Refugees. 2024.
16. Pavli A, Maltezos H. Health problems of newly arrived migrants and refugees in Europe. *J Travel Med*. 2017; 24(4).
17. Odone A, Tillmann T, Sandgren A, Williams G, Rechel B, Ingleby D, et al. Tuberculosis among migrant populations in the European Union and the European Economic Area. *Eur J Public Health*. 2015; 25(3): 506-12.
18. Figueroa-Munoz JI, Ramon-Pardo P. Tuberculosis control in vulnerable groups. *Bull World Health Organ*. 2008; 86(9): 733-5.
19. Pareek M, Greenaway C, Noori T, Munoz J, Zenner D. The impact of migration on tuberculosis epidemiology and control in high-income countries: a review. *BMC medicine*. 2016; 14(1): 48.
20. Rendon A, Centis R, Zellweger JP, Solovic I, Torres Duque C, Cordeiro CR, et al. Migration, TB control and elimination: whom to screen and treat. *Pulmonology*. 2018; 24(2): 99-105.
21. Meaza A, Tola HH, Eshetu K, Mindaye T, Medhin G, Gumi B. Tuberculosis among refugees and migrant populations: systematic review. *PLoS One*. 2022; 17(6): e0268696.
22. Ismail MB, Rafei R, Dabboussi F, Hamze M. Tuberculosis, war, and refugees: spotlight on the Syrian humanitarian crisis. *PLoS Pathog*. 2018; 14(6): e1007014.
23. Burki T. Infectious diseases in Malian and Syrian conflicts. *Lancet Infect Dis*. 2013; 13(4): 296-7.
24. Cousins S. Experts sound alarm as Syrian crisis fuels spread of tuberculosis. *BMJ*. 2014; 349: g7397.
25. Asadi L, Heffernan C, Menzies D, Long R. Effectiveness of Canada's tuberculosis surveillance strategy in identifying immigrants at risk of developing and transmitting tuberculosis: a population-based retrospective cohort study. *Lancet Public Health*. 2017; 2(10): e450-7.
26. van den Boogaard J, Slump E, Schimmel HJ, van der Hoek W, van den Hof S, de Vries G. High incidence of active tuberculosis in asylum seekers from Eritrea and Somalia in the first 5 years after arrival in the Netherlands. *Emerg Infect Dis*. 2020; 26(4): 675-81.
27. Vasiliu A, Köhler N, Altpeter E, Ægisdóttir TR, Amerali M, de Oñate WA, et al. Tuberculosis incidence in foreign-born people residing in European countries in 2020. *Euro Surveill*. 2023; 28(42): 2300051.
28. Stoycheva K, Cristea V, Ködmön C, Rosales-Klitz S, Zenner D, Vasiliu A, et al. Tuberculosis in people of Ukrainian origin in the European Union and the European Economic Area, 2019 to 2022. *Euro Surveill*. 2024; 29(12): 2400094.
29. Sanz B, Blasco T; ATBIM Project. Variables associated with diagnostic delay in immigrant groups with tuberculosis in Madrid. *Int J Tuberc Lung Dis*. 2007; 11(6): 639-46.
30. Quattrocchi A, Barchitta M, Nobile CGA, Prato R, Sotgiu G, Casuccio A, et al. Determinants of patient and health system delay among Italian and foreign-born patients with pulmonary tuberculosis: a multicentre cross-sectional study. *BMJ Open*. 2018; 8(8): e019673.
31. Holden IK, Lillebaek T, Andersen PH, Bjerrum S, Wejse C, Johansen IS. Extrapulmonary tuberculosis in Denmark from 2009 to 2014; characteristics and predictors for treatment outcome. *Open Forum Infect Dis*. 2019; 6(10): ofz388.
32. Zão I, Ribeiro AI, Apolinário D, Duarte R. Why does it take so long? The reasons behind tuberculosis treatment delay in Portugal. *Pulmonology*. 2019; 25(4): 215-22.
33. Kunst H, Lange B, Hovardovska O, Bockey A, Zenner D, Andersen AB, et al. Tuberculosis in adult migrants in Europe: a TB net consensus statement. *Eur Respir J*. 2025; 65(3): 2401612.
34. Hargreaves S, Lönnroth K, Nellums LB, Olaru ID, Nathavitharana RR, Norredam M, et al. Multidrug-resistant tuberculosis and migration to Europe. *Clin Microbiol Infect*. 2017; 23(3): 141-6.
35. Seung KJ, Keshavjee S, Rich ML. Multidrug-resistant tuberculosis and extensively drug-resistant tuberculosis. *Cold Spring Harb Perspect Med*. 2015; 5(9): a017863.
36. Walls T, Shingadia D. The epidemiology of tuberculosis in Europe. *Arch Dis Child*. 2007; 92(8): 726-9.
37. Hemming S, Windish P, Hall J, Story A, Lipman M. Treating TB patients with no entitlement to social support-welcome to the social jungle. *Thorax*. 2010; 65: A146.
38. Lillebaek T, Andersen AB, Dirksen A, Smith E, Skovgaard LT, Kok-Jensen A. Persistent high incidence of tuberculosis in immigrants in a low-incidence country. *Emerg Infect Dis*. 2002; 8(7): 679-84.
39. Klinkenberg E, Manissero D, Semenza JC, Verver S. Migrant tuberculosis screening in the EU/EEA: yield, coverage and limitations. *Eur Respir J*. 2009; 34(5): 1180-9.
40. Tavares AM, Fronteira I, Couto I, Machado D, Viveiros M, Abecasis AB, et al. HIV and tuberculosis co-infection among migrants in Europe: a systematic review on the prevalence, incidence and mortality. *PLoS One*. 2017; 12(9): e0185526.

41. Turkish Republic of North Cyprus, Ministry of Health.
42. Rendon A, Centis R, Zellweger JP, Solovic I, Torres-Duque CA, RobaloCordeiro C, et al. Migration, TB control and elimination: whom to screen and treat. *Pulmonology*. 2018; 24(2): 99-105.
43. Derendinger B, Mochizuki TK, Marcelo D, Shankar D, Mangeni W, Nguyen H, et al. C-reactive protein-based screening of people with tuberculosis symptoms: a diagnostic accuracy study. *Am J Respir Crit Care Med*. 2025; 211(3): 499-506.

Evaluation of Depressive Symptoms in Children Aged 3-6 Years: A Parent-Report Study

✉ Mine Cengiz¹, ✉ Yusuf Aktepe², ✉ Burcu Demir Gökmen²

¹Department of Medical Services and Techniques, Atatürk University Vocational Faculty of Health Services, Erzurum, Türkiye

²Department of Nursing, Ağrı İbrahim Çeçen University Faculty of Health Sciences, Ağrı, Türkiye

Abstract

BACKGROUND/AIMS: This study aimed to evaluate parent-reported depressive symptoms in children aged 3-6 years and to examine the relationship between these symptoms and the socio-demographic characteristics of the children and their parents.

MATERIALS AND METHODS: A descriptive cross-sectional study was conducted from January to March 2025 and included 236 parents who were recruited by simple random sampling. Data were collected using a socio-demographic information form and the child depressive symptom scale. The results were summarized as counts, percentages, means, and standard deviations, and comparisons were made using Independent Samples t-tests, one-way analysis of variance, Kruskal-Wallis, and Mann-Whitney U tests.

RESULTS: Depressive symptom levels varied significantly among children, differing by age, sex, and child health-related factors, such as allergic or chronic diseases. In addition, family-related factors, particularly parental chronic illness, were associated with higher levels of depressive symptoms ($p < 0.05$).

CONCLUSION: Children's depressive symptom levels differ according to certain child and family characteristics. The findings suggest that, when assessing the psychosocial well-being of young children, family-related characteristics, in addition to child-related factors, may also be taken into consideration. Further studies with larger sample sizes, longitudinal designs, and intervention-based approaches are needed to evaluate the potential benefits of psychoeducation and support programs for parents.

Keywords: Parents, child, preschool, depression, symptoms

INTRODUCTION

Raising children is one of the primary responsibilities of adults across all cultures. Parents are among the primary factors in children's social and emotional development and can adopt authoritarian, permissive, or democratic approaches during this process. These parenting styles may have positive or negative impacts on children's personalities, mental constructs, and emotional states.¹ Many children today do not receive adequate attention because of nuclear family structures, double-income households, and other problems. This decreased attention can affect children's mental health. Anxiety and depressive symptoms may be more common in children.²

It is increasingly reported, both globally and nationally, that mental illnesses have an earlier onset than previously believed.^{3,4} The 2022 mental health report published by the World Health Organization (WHO) indicates that depression in preschool-aged children remains insufficiently researched⁴, whereas research focusing on depression and anxiety in school-age children and adolescents has reported an estimated prevalence of 4-5%.³ Although significant progress has been made in recent years toward understanding depression in school-aged children and adolescents, research into the prevalence of depression in children under 6 years of age has historically been scarce, and the field remains nascent.⁵ The existing data on this topic are also highly

To cite this article: Cengiz M, Aktepe Y, Demir Gökmen B. Evaluation of depressive symptoms in children aged 3-6 years: a parent-report study. Cyprus J Med Sci. 2026;11(5):342-352

ORCID IDs of the authors: M.C. 0000-0002-2066-3464; Y.A. 0009-0004-6815-0387; B.D.G. 0000-0003-2058-8924.



Corresponding author: Burcu Demir Gökmen

E-mail: burcudmr04@gmail.com

ORCID ID: orcid.org/0000-0003-2058-8924

Received: 09.12.2025

Accepted: 22.05.2026

Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.

This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

variable. A meta-analysis reported that the prevalence of depression among children and adolescents ranged from 1% to 58%.⁶ This broad range may be attributed to measurement tools, sample characteristics, and cultural differences.

Studies focusing on the preschool period report that the prevalence of depression typically ranges from 0.5% to 2%, with some community-based studies indicating rates as high as 2-5%.^{7,8} Although these rates appear lower than those observed during adolescence, early-onset depression remains a clinically significant issue because it carries a risk of persistence.⁶

Recent studies indicate that depression is more prevalent among girls than boys.^{9,10} Furthermore, low and middle socio-economic statuses are identified as significant risk factors. Because a large proportion of children and adolescents worldwide reside in low- and middle-income countries, a substantial number of children and adolescents may be affected by depression. Exposure to adverse living conditions, such as parental migration, violence, conflict, and poverty, has also been reported to increase the risk of depression. Additionally, the COVID-19 pandemic has reportedly exacerbated these existing challenges, leading to increased levels of depression in children and adolescents.^{11,12} Finally, a substantial body of research demonstrates that depressive symptoms emerging during the preschool period serve as a strong risk indicators for other psychopathologies that may develop later in life. Therefore, it is essential to recognize and monitor early-onset depressive symptoms and plan appropriate interventions.^{12,13}

Depression can develop in preschool-age children due to early adverse environments and ongoing, acute, and chronic stressors. Depression in children is associated with symptoms such as aggression, separation anxiety, impulsivity, hyperactivity, social adjustment difficulties, and impaired cognition.¹⁴ A review of the literature indicates that both international and national studies report associations between depressive symptoms in preschool children and a range of mental disorders and symptoms, including sleep problems, irritability, separation anxiety, childhood traumas, anxiety disorders, behavioral disorders, hyperactivity disorder, aggression, cognitive impairment, excessive guilt, sadness, motor inhibition, and attention deficit.^{15,16}

Due to the limited capacity of preschool children to verbally express their emotions, depressive symptoms are predominantly observed through behavioral, emotional, and physical changes. The literature indicates that, although clinical experts typically employ diagnostic scales to assess depressive symptoms in this age group, parents, teachers, and caregivers also provide crucial information.^{12,16,17} Alongside traditional assessment methodologies, emerging digital approaches, including machine learning-based analysis of children's speech, have shown promise in detecting anxiety and depressive symptoms during early childhood.¹⁸ Although professional expertise is required for clinical assessment, parents are recognized as the primary source of information during early childhood, as they observe the child's daily behavior over extended periods and in the most natural environments.¹⁹

Among the depressive symptoms most frequently reported by families are persistent sadness or dysphoria in the child, frequent crying, and diminished interest in play and activities they previously enjoyed. Other notable behavioral changes include social withdrawal, a preference for isolation, excessive irritability, and outbursts of anger.²⁰ While some children may become excessively dependent on their parents,

others may avoid social interactions. Changes in sleep and appetite patterns, and somatic symptoms such as unexplained headaches or abdominal pain, are also significant indicators of depression.²¹ In their study examining parent-reported data, Wakschlag et al.²² emphasized that while the low-frequency occurrence of certain behaviors may be considered developmentally normal, an increase in the frequency and severity of these behaviors can serve as a risk indicator for psychopathology. Conversely, certain behaviors that occur infrequently but with high intensity (e.g., expressions of profound sadness severe enough to interfere with a child's eating) should be directly evaluated as indicators of clinical risk.²²

Consequently, early recognition of these symptoms by families is critical to the child's mental health. Early awareness facilitates the timely initiation of appropriate psychosocial interventions and can prevent progression. By mitigating the risk of chronicity, early intervention can also avert subsequent psychiatric issues that the child may encounter later in life.^{12,20} Furthermore, given the potential for depression to adversely affect the child's social, emotional, and cognitive development, early diagnosis is important for preventing developmental deficits. Therefore, rather than dismissing these symptoms as "normal," "spoiled behavior," or "a passing phase," families should seek professional support to promote healthy development. Furthermore, a 2024 joint report by the WHO and the United Nations Children's Fund draws attention to the rising incidence of mental health issues, such as depression, among children and underscores the critical need for early support and intervention.²³

Although depression is less frequently observed during the preschool period, it must be addressed as a significant mental health issue, necessitating early detection and appropriate intervention.^{6,23} The present study, using parent-reported data, was conducted to identify depressive symptoms and determine associated risk factors in children aged 3 to 6 years.

Research Questions:

- Are depressive symptoms present in children aged 3-6 years according to parent report?
- Is there a difference in parent-reported depressive symptoms in children aged 3-6 years according to children's socio-demographic characteristics (age, sex, academic performance, allergic disease, chronic disease/medication use)?
- Is there a difference in parent-reported depressive symptoms in children aged 3-6 years according to parents' socio-demographic characteristics (age, sex, marital status, education level, income level, chronic disease, history of depression or other mental disorder)?

MATERIALS AND METHODS

Study Design and Setting

This descriptive study was conducted from January to March 2025 in the central district of Ağrı Province, Türkiye. Eighteen neighborhoods in the central district of Ağrı Province were initially divided into three strata according to socio-economic status. Among the six neighborhoods with the highest socio-economic status, one was selected by lottery. The study population comprised the parents of 518 children aged 3-6 years who attended three preschools in this neighborhood. The schools' names

and student rosters were obtained with permission from the provincial Directorate of National Education. The study sample consisted of parents of students from the three target preschools; the parents were selected by simple random sampling (lottery method).

Sample size was calculated for a known population using the formula $n = (N \cdot t^2 \cdot p \cdot q) / (d^2 \cdot (N - 1) + (t^2 \cdot p \cdot q))$.²⁴ The minimum sample size was determined to be 221 participants. The study included 236 parents who agreed to participate and completed the data collection forms. The study included Turkish-speaking parents aged 18 years or older who had a child between 3 and 6 years of age and had no sensory impairments that would hinder communication.

Data Collection Forms

The study data were collected using a socio-demographic information form and the child depressive symptom scale (CDSS).

Socio-Demographic Information Form

This form was developed by the researchers based on a review of the relevant literature.^{25,26} It included 14 questions about the socio-demographic characteristics of the participating parent (age, gender, marital status, education level, subjective income level, presence of chronic disease, history of depression, history of psychiatric illness) and their child (age, sex, school attended, performance at school, presence of allergic disease, presence of chronic disease, and regular medication use).

Child Depressive Symptom Scale

The CDSS, which was developed by Erol et al.²⁷ in 2020 as a broad measure of childhood problems, consists of 56 items in 8 subscales: social adaptation (items 16, 18, 19, 21, 22, 23, 31, 45, 50, and 54), impulsivity (items 9, 10, 28, 30, 44, and 51), separation anxiety (items 1, 5, 15, 17, 37, and 38), impaired cognitive functioning (items 2, 3, 6, 7, 11, 12, 32, 34, 36, 48, and 53), somatization (items 4, 8, 13, 26, and 39), fantasy-reality confusion (items 14, 20, 27, 29, 33, 46, 49, and 52), archaic anxiety (items 40, 43, and 47), and aggression/destructiveness (items 24, 25, 35, 41, 42, 55, and 56). Parents complete the CDSS. Items are scored on a 5-point Likert-type scale from 1 (never) to 5 (always); the social adaptation subscale is reverse-scored. Scale results can be evaluated according to the total score (possible range: 56-280) and the subscale scores. Higher scores indicate a greater tendency toward depression in children.

The Cronbach's alpha coefficients for the scale were reported as follows: 0.95 for the total scale; 0.85 for the social adaptation subscale; 0.82 for impulsivity; 0.79 for separation anxiety; 0.91 for impaired cognitive functioning; 0.75 for somatization; 0.84 for fantasy-reality confusion; 0.71 for archaic anxiety; and 0.90 for aggression/destructiveness.²⁷ In the present study, the Cronbach's alpha coefficients were calculated as follows: 0.88 for the total scale; 0.71 for social adaptation; 0.64 for impulsivity; 0.78 for separation anxiety; 0.54 for impaired cognitive functioning; 0.75 for somatization; 0.52 for fantasy-reality confusion; 0.82 for archaic anxiety; and 0.83 for aggression/destructiveness. The relatively low internal consistency observed for the impaired cognitive functioning and fantasy-reality confusion subscales may be attributed to several factors. These constructs may not yet be fully manifest during the preschool period, and parents have limited insight into their children's internal cognitive processes. Furthermore, variance may be low because these specific symptoms are infrequent in the sample.

Data Collection

Preliminary meetings were held with parents to provide information about the study. These meetings took place in a room designated by the school administration during the students' morning drop-off or afternoon pick-up times. Following these briefings, verbal and written informed consent was obtained from the participating parents. Data were then collected by one of two methods, depending on participant preference. Parents could complete the questionnaires either immediately in the meeting room (average completion time 25 to 30 minutes) or later at home. Parents who preferred completing the forms at home were provided with instructions and the forms, which they returned during the same week.

Ethical Considerations

All procedures involving human participants in this study were conducted in accordance with the ethical standards of the institutional research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Approved the study protocol Ağrı İbrahim Çeçen University Scientific Research Ethics Committee (approval number: 406, date: 28.11.2024), and written permission was obtained from the institution where the research was conducted (date: 12.12.2024, application no: MEB. TT.2024.011071). Permission to use the CDSS was also obtained from the creator of the scale. All participants provided written informed consent.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, version 22.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics were presented as frequencies, percentages, means, standard deviations (SDs), and ranges. The normality of data distributions was assessed using skewness-kurtosis values and histograms. Continuous variables showing a normal distribution were analyzed using parametric tests (Independent Samples t-tests for comparisons of two groups; one-way analysis of variance (ANOVA) for comparisons of three or more groups). Variables that did not meet the assumption of normality were analyzed using non-parametric tests (Mann-Whitney U test or Kruskal-Wallis test). Among the main outcome measures, the CDSS total score and the social adaptation, impulsivity, separation anxiety, impaired cognitive functioning, somatization, fantasy-reality confusion, and archaic anxiety subscale scores were normally distributed, while aggression/destructiveness scores were not normally distributed. $p < 0.05$ was used as the threshold for statistical significance. Effect sizes were also determined to demonstrate the clinical relevance of statistically significant results. Effect sizes were calculated: Cohen's d for Independent Samples t-tests, eta squared (η^2) for one-way ANOVA, and r for non-parametric tests (Mann-Whitney U and Kruskal-Wallis). In addition, Cronbach's alpha values were examined to assess the reliability of the scale.

RESULTS

The socio-demographic characteristics of the participating parents and their children are shown in Table 1. The mean age of the parents in the study was 33.50 (SD =4.88) years. Among the participants, 73.3% were mothers, 85.2% were married, 67.4% had university-level or higher education, and 63.6% reported a high subjective income level (income covering expenses). Most parents had no chronic disease (77.5%), no family history of depression (93.2%), and no family history of psychiatric

illness (95.8%). Among their children, 63.1% were girls and 34.7% were 5 years old. Most children had neither allergic disease (90.3%) nor chronic illness/regular medication use (97%) (Table 1).

The mean total and subscale scores reported by parents on the CDSS are shown in Table 2. The mean total score of the study participants was 123.69 (SD =19.67). Their mean subscale scores were 37.35 (SD =9.10)

on the social adaptation subscale, 14.46 (SD =4.63) on the impulsivity subscale, 13.09 (SD =4.17) on the separation anxiety subscale, 18.80 (SD =5.98) on the impaired cognitive functioning subscale, 8.54 (SD =2.97) on the somatization subscale, 13.75 (SD =5.00) on the fantasy-reality confusion subscale, 6.51 (SD =2.81) on the archaic anxiety subscale, and 11.18 (SD =4.97) on the aggression/destructiveness subscale (Table 2).

Table 1. Socio-demographic characteristics of the participating parents and their children (n=236)

			Mean (SD)	Range
Age (years)			33.50 (4.88)	24-49
			n	%
Parental socio-demographic characteristics	Gender	Mother	173	73.3
		Father	63	26.7
	Marital status	Married	201	85.2
		Single	35	14.8
	Education level	Elementary/middle school	35	14.8
		High school	42	17.8
		University or higher	159	67.4
	Subjective monthly income	Income covers expenses	150	63.6
		Income partially covers expenses	50	21.2
		Income inadequate compared to expenses	36	15.3
Child socio-demographic characteristics	Chronic disease		53	22.5
	Family history of depression		16	6.8
	Family history of psychiatric Ill ness		10	4.2
	Age (years)	3	36	15.3
		4	57	24.2
		5	82	34.7
		6	61	25.8
	Sex	Female	149	63.1
		Male	87	36.9
	Academic performance	Good	108	45.8
		Fair	112	47.5
		Poor	16	6.7
	Allergic disease		23	9.7
	Chronic disease and regular medication use		7	3.0

n: Number, SD: Standard deviation.

Table 2. Parents' mean total and subscale scores on the CDSS

CDSS				
Subscale	Number of items	Possible range	Observed range	Mean (SD)
Social adaptation	10	10-50	10-50	37.35 (9.10)
Impulsivity	6	6-30	6-30	14.46 (4.63)
Separation anxiety	6	6-30	6-28	13.09 (4.17)
Impaired cognitive functioning	11	11-55	11-42	18.80 (5.98)
Somatization	5	5-25	5-18	8.54 (2.97)
Fantasy-reality confusion	8	8-40	8-33	13.75 (5.00)
Archaic anxiety	3	3-15	3-15	6.512 (2.81)
Aggression/destructiveness	7	7-35	7-32	11.18 (4.97)
Total	56	56-280	56-191	123.69 (19.67)

CDSS: Child depressive symptom scale, SD: Standard deviation.

Comparisons of mean CDSS total scores according to the parents' socio-demographic characteristics are shown in Table 3. Parents with chronic disease had higher mean scores than those without chronic disease ($p<0.05$). The magnitude of this difference was moderate (Cohen's $d=0.59$), indicating a clinically meaningful association. However, there were no statistically significant differences in mean CDSS total scores by gender, marital status, education level, subjective income level, or family history of depression and psychiatric illness ($p>0.05$) (Table 3).

In Table 4, mean CDSS total scores are compared across the socio-demographic characteristics of the participants' children. Mean scores were higher for children with poor school performance compared to those with fair or good performance ($p<0.05$; $\eta^2=0.03$) and for children with allergic or other chronic diseases who used regular medication compared to those without such conditions or medication use ($p<0.05$; $r=0.14$). There were no statistically significant differences with respect to the children's age, sex, or school attended ($p>0.05$) (Table 4).

Comparisons of CDSS subscale scores according to the parents' socio-demographic characteristics are shown in Table 5. Somatization subscale scores were higher in mothers than in fathers ($p=0.05$). Parental chronic disease was associated with significantly higher subscale scores for separation anxiety ($d=0.37$), impaired cognitive functioning ($d=0.55$),

somatization ($d=0.52$), fantasy-reality confusion ($d=0.65$), archaic anxiety ($d=0.32$), and aggression/destructiveness ($d=0.70$) ($p<0.05$) (Table 5).

Comparisons of CDSS subscale scores by children's socio-demographic characteristics are presented in Table 6. When compared across ages, fantasy-reality Confusion ($\eta^2=0.04$) subscale scores were higher in 5-year-olds than in 3-, 4-, or 6-year-olds, whereas aggression/destructiveness ($r=0.19$) subscale scores were higher in 6-year-olds than in children of other ages ($p<0.05$). In sex-based comparisons, impulsivity ($d=0.44$) subscale scores were higher for boys than for girls, while archaic anxiety ($d=0.27$) subscale scores were higher for girls than for boys ($p<0.05$) (Table 6).

Poor academic performance, as reported by the parents, was associated with higher Impulsivity ($\eta^2=0.04$), separation anxiety ($\eta^2=0.05$), impaired cognitive functioning ($\eta^2=0.04$), and archaic anxiety ($\eta^2=0.04$) subscale scores than those for fair and good academic performance ($p<0.05$) (Table 6).

Separation anxiety and somatization subscale scores were higher in children with allergic disease ($p<0.05$; $r=0.13$ and $r=0.22$, respectively) and in those with any other chronic disease or regular medication use ($p<0.05$; $r=0.17$ and $r=0.15$, respectively) compared to children without these conditions. In contrast, aggression/destructiveness subscale scores were higher in children without allergic disease than in those with allergic disease ($p<0.05$; $r=0.15$) (Table 6).

Table 3. Comparison of the mean CDSS scores according to parental socio-demographic characteristics

Socio-demographic characteristics	CDSS	
	Mean (SD)	Test/p-value
Gender		
Mother	123.75 (20.12)	$t=0.076$
Father	123.52 (18.52)	$p=0.939$
Marital status		
Married	124.47 (17.46)	$t=1.039$
Single	119.17 (29.28)	$p=0.305$
Education level		
Elementary-middle school	124.71 (19.90)	$\chi^2=5.194$
High school	123.21 (18.09)	$p=0.268$
University or higher	122.88 (19.87)	
Subjective monthly income		
Income covers expenses	124.37 (18.38)	$F=1.211$
Income partially covers expenses	124.98 (20.23)	$p=0.300$
Income inadequate compared to expenses	119.03 (23.67)	
Chronic disease		
Yes	132.43 (14.62)	$t=3.779$
No	121.15 (20.24)	$p<0.001^*$ $d=0.59$
Family history of depression		
No	123.42 (19.45)	$U=1663.0$
Yes	127.38 (22.86)	$p=0.713$
Family history of psychiatric illness		
No	123.54 (19.78)	$U=1043.0$
Yes	127.10 (17.52)	$p=0.680$

* $p<0.05$; CDSS: Child depressive symptom scale, n: Number, U: Mann-Whitney U, F: One-way ANOVA, p: Significance value, t: Independent Samples t-test, χ^2 : Chi-square, d: Cohen's d, ANOVA: Analysis of variance.

Table 4. Comparison of mean CDSS scores according to the socio-demographic characteristics of the participants' children

Socio-demographic characteristics	CDSS	
	Mean (SD)	Test/p-value
Age (years)		
3	121.36 (23.99)	$F=0.563$
4	121.79 (14.84)	$p=0.640$
5	125.04 (19.91)	
6	125.02 (20.70)	
Gender		
Female	123.11 (20.24)	$t=-0.584$
Male	124.67 (18.73)	$p=0.560$
Performance in school		
Good	124.16 (19.83)	$F=3.207$
Fair	122.83 (18.89)	$p=0.024^*$
Poor	143.00 (24.39)	$\eta^2=0.03$
Allergic disease		
No	122.82 (19.16)	$U=1847.500$
Yes	131.70 (22.82)	$p=0.050^*$
Chronic disease and regular medication use		
No	123.03 (19.09)	$U=427.000$
Yes	145.29 (27.42)	$p=0.035^*$ $r=0.14$

* $p<0.05$; CDSS: Child depressive symptom assessment scale, n: Number, U: Mann-Whitney U, F: One-way ANOVA, p: Significance value, t: Independent Samples t-test, η^2 : Eta squared, r: Effect size for non-parametric test, ANOVA: Analysis of variance.

Table 5. Mean CDSS subscale scores according to parental socio-demographic characteristics								
Socio-demographic characteristics	Social adaptation mean (SD)	Impulsivity mean (SD)	Separation anxiety mean (SD)	Impaired cognitive functioning mean (SD)	Somatization mean (SD)	Fantasy-reality confusion mean (SD)	Archaic anxiety mean (SD)	Aggression/destructiveness mean (SD)
Gender								
Mother	36.94 (9.044)	14.47 (4.83)	13.15 (4.24)	19.04 (6.07)	8.76 (3.00)	13.77 (4.78)	6.53 (2.82)	11.08 (4.93)
Father	38.46 (9.22)	14.43 (4.07)	12.92 (3.98)	18.16 (5.72)	7.92 (2.80)	13.71 (5.58)	6.46 (2.80)	11.46 (5.11)
Test	t=-1.135	t=0.066	t=0.374	t=0.996	t=1.942	t=0.074	t=0.173	U=5110.500
p-value	p=0.258	p=0.947	p=0.709	p=0.320	p=0.050*	p=0.941	p=0.863	p=0.460
Marital status								
Married	38.07 (8.67)	14.43 (4.41)	13.16 (4.09)	18.79 (5.89)	8.56 (2.94)	13.89 (5.00)	6.40 (2.63)	11.17 (4.74)
Single	33.20 (10.44)	14.63 (5.82)	12.66 (4.65)	18.89 (6.56)	8.40 (3.16)	13.00 (4.93)	7.17 (3.67)	11.23 (6.23)
Test	t=2.971	t=-0.190	t=0.663	t=-0.091	t=0.298	t=0.969	t=-1.196	U=3220.500
p-value	p=0.003	p=0.851	p=0.508	p=0.928	p=0.766	p=0.334	p=0.239	p=0.421
Education level								
Elementary-middle school	34.44 (7.74)	15.37 (4.58)	13.64 (4.64)	20.63 (7.09)	7.86 (2.69)	14.76 (6.02)	6.46 (2.26)	11.55 (4.39)
High school	38.86 (9.04)	14.62 (3.72)	11.71 (3.73)	17.76 (5.23)	9.26 (3.41)	13.95 (4.45)	6.64 (2.27)	10.40 (4.09)
University or higher	37.79 (8.75)	14.17 (4.79)	13.26 (4.16)	18.50 (5.55)	8.34 (2.73)	13.37 (5.07)	6.39 (2.97)	11.06 (4.99)
Test	X ² =7.647	X ² =3.486	X ² =5.699	X ² =6.727	X ² =3.942	X ² =6.127	X ² =4.279	X ² =8.217
p-value	p=0.105	p=0.480	p=0.223	p=0.151	p=0.414	p=0.190	p=0.370	p=0.084
Subjective monthly income								
Income covers expenses	37.51 (9.11)	14.63 (4.83)	13.37 (4.19)	19.03 (5.89)	8.54 (2.82)	13.65 (4.61)	6.35 (2.67)	11.29 (5.03)
Income partially covers expenses	36.80 (8.77)	14.44 (4.13)	13.02 (4.12)	19.24 (6.01)	8.78 (3.23)	14.90 (5.19)	6.70 (3.04)	11.10 (5.04)
Income inadequate compared to expenses	37.42 (9.69)	13.81 (4.53)	12.03 (4.09)	17.25 (6.24)	8.19 (3.23)	12.58 (5.97)	6.92 (3.06)	10.83 (4.72)
Test	F=0.116	F=0.455	F=1.513	F=1.459	F=0.406	F=2.366	F=0.724	X ² =0.165
p-value	p=0.891	p=0.635	p=0.222	p=0.235	p=0.667	p=0.096	p=0.486	p=0.921
Chronic disease								
Yes	35.51 (9.19)	14.49 (3.70)	14.26 (4.25)	21.28 (6.21)	9.70 (3.05)	16.21 (5.94)	7.21 (2.89)	13.77 (5.74)
No	37.88 (9.03)	14.45 (4.88)	12.75 (4.09)	18.08 (5.73)	8.20 (2.86)	13.04 (4.45)	6.31 (2.76)	10.43 (4.47)
Test	t=-1.677	t=0.059	t=2.353	t=3.514	t=3.301	t=3.598	t=2.060	U=3105.000
p-value	p=0.095	p=0.953	p=0.019*	p=0.001*	p=0.001*	p=0.001*	p=0.041*	p=0.000*
d			d=0.37	d=0.55	d=0.52	d=0.65	d=0.32	d=0.70
Family history of depression								
No	37.21 (9.25)	14.51 (4.64)	12.97 (4.16)	18.78 (5.93)	8.54 (3.00)	13.86 (4.95)	6.40 (2.65)	11.15 (4.84)
Yes	39.19 (6.65)	13.75 (4.67)	14.75 (4.06)	19.06 (6.78)	8.56 (2.61)	12.31 (5.47)	8.13 (4.21)	11.63 (6.66)
Test	U=1631.500	U=1582.000	U=1319.000	U=1754.000	U=1682.500	U=1347.000	U=1399.500	U=1666.500
p-value	p=0.626	p=0.499	p=0.093	p=0.982	p=0.767	p=0.116	p=0.168	p=0.720
Family history of psychiatric illness								
No	37.39 (9.26)	14.39 (4.61)	12.99 (4.15)	18.81 (5.99)	8.53 (3.00)	13.81 (5.04)	6.51 (2.77)	11.09 (4.95)
Yes	36.50 (4.12)	16.00 (5.23)	15.30 (4.17)	18.50 (6.13)	8.70 (2.16)	12.40 (3.57)	6.50 (3.75)	13.20 (5.33)
Test	U=925.000	U=917.000	U=761.000	U=1087.500	U=990.000	U=992.500	U=1031.500	U=856.000
p-value	p=0.331	p=0.312	p=0.080	p=0.840	p=0.503	p=0.514	p=0.638	p=0.190
*p<0.05; CDSS: Child depressive symptom scale, U: Mann-Whitney U, F: One-way ANOVA, p: Significance value, t: Independent Samples t-test, X ² : Chi-square; *d: Cohen's d, SD: Standard deviation, ANOVA: Analysis of variance.								

Table 6. Mean CDSS subscale scores according to the socio-demographic characteristics of the participants' children

Socio-demographic characteristics	Social adaptation mean (SD)	Impulsivity mean (SD)	Separation anxiety Mean (SD)	Impaired cognitive functioning mean (SD)	Somatization mean (SD)	Fantasy-reality confusion mean (SD)	Archaic anxiety mean (SD)	Aggression/ destructiveness mean (SD)
Child's age (years)								
3	39.42 (9.49)	14.14 (5.17)	13.22 (4.60)	17.47 (5.85)	7.78 (3.15)	12.67 (5.32)	6.56 (3.26)	10.11 (5.19)
4	37.11 (8.47)	15.51 (4.49)	13.30 (4.24)	18.53 (5.82)	8.39 (2.71)	12.42 (3.67)	6.12 (2.36)	10.42 (3.44)
5	37.98 (9.41)	13.54 (4.13)	12.81 (3.78)	19.43 (5.99)	8.89 (3.18)	14.71 (5.34)	6.48 (2.64)	11.22 (5.18)
6	35.51 (8.86)	14.92 (4.91)	13.20 (4.42)	19.00 (6.19)	8.66 (2.75)	14.36 (5.10)	6.90 (3.12)	12.48 (5.55)
Test	F=1.608	F=2.356	F=0.199	F=0.954	F=1.260	F=3.318	F=0.762	X ² = 8.170
p- value	p=0.188	p=0.073	p=0.897	p=0.415	p=0.289	p=0.021* η ² =0.04	p=0.517	p=0.043* r=0.19
Child's sex								
Female	37.05 (9.77)	13.72 (4.39)	13.08 (4.13)	19.09 (5.96)	8.58 (2.98)	13.64 (4.96)	6.79 (2.82)	11.18 (4.82)
Male	37.86 (7.83)	15.74 (4.79)	13.10 (4.26)	18.31 (6.02)	8.47 (2.95)	13.95 (5.07)	6.03 (2.73)	11.20 (5.24)
Test	t=-.703	t=-3.294	t=-.041	t=0.963	t=0.264	t=-0.469	t=2.012	U=6446.500
p- value	p=0.483	p=0.001* d=0.44	p=0.968	p=0.337	p=0.792	p=0.639	p=0.045* d=0.27	p=0.944
Performance in school								
Good	38.38 (9.88)	13.71 (4.07)	12.55 (3.54)	18.09 (6.14)	8.49 (2.88)	14.24 (5.66)	7.23 (2.95)	11.46 (5.60)
Fair	36.50 (8.43)	14.88 (4.89)	13.55 (4.40)	19.12 (5.46)	8.55 (2.96)	13.27 (4.24)	5.95 (2.55)	11.01 (4.27)
Poor	33.54 (8.11)	19.54 (6.92)	17.79 (5.45)	25.93 (4.17)	10.15 (5.15)	16.07 (3.90)	6.65 (1.99)	14.40 (7.58)
Test	F=1.583	F=4.222	F=5.852	F=4.166	F=0.952	F=1.513	F=5.222	X ² =2.795
p- value	p=0.194	p=0.006* η ² =0.04	p=0.001* η ² =0.05	p=0.007* η ² =0.04	p=0.416	p=0.212	p=0.002* η ² =0.04	p=0.424
Allergic disease in child								
No	37.20 (9.30)	14.39 (4.40)	12.87 (4.02)	18.63 (5.96)	8.13 (2.70)	13.79 (5.04)	6.44 (2.82)	11.37 (4.98)
Yes	38.70 (7.00)	15.09 (6.51)	15.13 (5.00)	20.35 (6.05)	12.35 (2.60)	13.44 (4.63)	7.22 (2.63)	9.44 (4.60)
Test	U=2311.500	U=2403.500	U=1841.000	U=2020.000	U=599.500	U=2395.000	U=1979.000	U=1740.000
p- value	p=0.657	p=0.882	p=0.050* r=0.13	p=0.167	p<0.001* r=0.22	p=0.860	p=0.127	p=0.021* r=0.15
Chronic disease and regular medication use								
No	37.39 (9.07)	14.38 (4.50)	12.94 (4.04)	18.67 (5.76)	8.43 (2.84)	13.66 (4.95)	6.50 (2.79)	11.07 (4.83)
Yes	36.00 (10.58)	17.29 (7.85)	18.14 (5.46)	23.14 (10.76)	12.14 (4.67)	16.71 (5.85)	7.00 (3.56)	14.86 (8.13)
Test	U=737.500	U=646.500	U=342.500	U=624.500	U=408.000	U=539.500	U=757.000	U=619.000
p- value	p=0.719	p=0.382	p=0.010* r=0.17	p=0.319	p=0.026* r=0.15	p=0.139	p=0.801	p=0.300

*p<0.05; CDSS: Child depressive symptom scale, U: Mann-Whitney U, F: One-way ANOVA, p: Significance value, t: Independent Samples t-test, X²: Chi-square, η²: Eta squared, d: Cohen's d, r: Effect size for non-parametric tests, SD: Standard deviation, ANOVA: Analysis of variance.

DISCUSSION

This study was conducted to evaluate depressive symptoms in 3- to 6-year-old children, as reported by their parents, and to identify factors associated with these symptoms. While research on depressive tendencies in children has been limited historically, it has recently received increased attention owing to growing awareness of mental health and its importance. However, the measurement tool used in this study was developed in 2020, and few studies have been conducted on the subject. Based on the reports of the parents participating in our study, the mean total score for children's depressive symptoms fell within the range of 56-191, which is below the midpoint of the possible score range (56-280).

Scores were lower for seven of the subscales (impulsivity, separation anxiety, impaired cognitive functioning, somatization, fantasy-reality confusion, archaic anxiety, and aggression/destructiveness) and were highest for the social adaptation subscale (Table 2).

However, the relatively low internal consistency coefficients observed for the impaired cognitive functioning and fantasy-reality confusion subscales (Cronbach's alpha =0.54 and 0.52, respectively) suggest compromised measurement reliability, which may limit the generalizability of the study findings.

Higher scores on the social adaptation subscale indicate that children are socially maladjusted. Consistent with our results, another study

conducted in a similar group reported low scores for impaired cognitive functioning, impulsivity, fantasy-reality confusion, archaic anxiety, aggression, and somatization.²⁸ However, the children in that study were reported to have better social adaptation than those in the current study. This may be because the study was conducted in a provincial center, where rural and urban life are intertwined, resulting in strong social relations and interpersonal interactions.²⁹ The social maladjustment observed in the present study may partly reflect changes in children's social interactions and communication patterns associated with increasing exposure to digital media and touchscreen technologies.³⁰

Our findings are similar to those of a study reporting that depressive symptom scores were higher for social adaptation but lower for impaired cognitive functioning, impulsivity, fantasy-reality confusion, archaic anxiety, aggression/destructiveness, and somatization.³¹

In this study, parents with chronic diseases reported higher levels of depressive symptoms in their children compared with parents without chronic diseases (Tables 3 and 5). This suggests that parental health problems may increase children's tendencies toward depressive symptoms. The stress, anxiety, and physical limitations experienced by parents while coping with their chronic disease may adversely impact the mental health of their children. This finding suggests that the psychological and physical stress experienced by parents while coping with chronic disease may be reflected in children's emotional and cognitive functioning. Parental chronic disease may lead to children having higher levels of anxiety (separation anxiety, archaic anxiety), disordered perception and cognition (impaired cognitive functioning, fantasy-reality confusion), more intense physical complaints (somatization), and occasional emotional dysregulation (aggression/destructiveness).^{32,33} Previous studies have reported that the children of parents with chronic disease have increased internalizing problems (anxious, depressive, and introverted behaviors, and somatic complaints) and externalizing problems (anger, aggression, impulsivity, disobedience, and uncooperativeness).^{32,33} These findings suggest that, considering parents' influence as role models, observing chronic disease in a parent may influence a child's mental health.

Children with poor school performance had higher depressive symptom scores overall and across symptom subscales (Tables 4 and 6). School performance that does not meet expectations may negatively affect children by leading to stress and low self-confidence.^{34,35} In an Australian study, depression and anxiety symptoms were found to lead to significant learning losses, especially in motor skills. The authors concluded that childhood anxiety and depression are not only predictors of later mental health problems but also adversely impact academic achievement.³⁶ In a study conducted in Taiwan, students with low academic performance had a higher risk of depressive symptoms.³⁷ Several studies conducted among young children have shown that emotional and behavioral difficulties, including aggressive behaviors and separation anxiety, are associated with poorer academic outcomes.^{38,39} Thus, the results of our study are consistent with those reported in the literature.

Children with allergic diseases had higher total depressive symptom scores and higher scores on separation anxiety and somatization subscales, whereas children without allergic diseases had higher aggression/destructiveness scores (Tables 4 and 6). Previous studies have shown that allergic diseases are associated with increased mental health problems and internalizing symptoms during childhood. These associations may be related to the psychological burden of chronic

disease, symptom uncertainty, and repeated healthcare experiences.^{40,41} However, the mechanisms underlying the relationship between allergic diseases and specific emotional symptoms remain unclear. Therefore, findings related to the somatization subscale should be interpreted cautiously due to its relatively low internal consistency in the present sample.

Similarly, the presence of chronic disease and regular medication use in the child were associated with higher total depressive symptom scores and higher subscale scores for separation anxiety and somatization (Tables 4 and 6). This is consistent with a 2011 meta-analysis by Pinquart and Shen that showed that children and adolescents with chronic physical illness exhibited higher levels of depressive symptoms than healthy peers. Conditions such as chronic fatigue syndrome, fibromyalgia, migraine, and, in particular, epilepsy were associated with more pronounced effects.⁴² Chronic health problems may adversely affect children's mental health and put them at greater risk of psychological consequences, because they have difficulty coping with physical disorders and because they face associated limitations on their social lives. Other recent studies reported that children and adolescents with chronic diseases have a higher incidence of anxiety and depression than their healthy peers.^{43,44} In the literature, chronic diseases and long-term medication use in preschool students have been suggested to hinder motor, social, and cognitive development due to physical restrictions, low energy levels, perception of the disease as punishment, and development of immature behaviors.⁴²

Regarding gender, mothers in our study scored significantly higher on the somatization subscale of the CDSS (Table 5), suggesting that mothers may be more aware of children's physical complaints than fathers. Considering the association between somatic symptoms and anxiety, the greater importance mothers place on their children's physical complaints may reflect greater maternal awareness of anxiety-related behaviors. Similarly the study reported that mothers placed greater importance on their children's anxious behaviors.⁴⁵ However, our findings should be interpreted in light of the low internal consistency coefficients obtained for the separation anxiety and somatization subscales, as well as the small number of children with chronic disease and regular medication use.

Scores on the fantasy-reality confusion subscale were highest for 5-year-old children in this study (Table 6). There is no research directly related to this finding in the literature. However, the perception of objective reality develops with age, and there may be age-related differences in understanding certain concepts. In addition, children's ability to distinguish between what is real and what is not is not sufficiently developed before starting kindergarten.⁴⁶

Scores in the aggression/destructiveness subscale were highest for the 6-year-old children in this study (Table 6). The impending transition from kindergarten to elementary school, along with the changes 6-year-old children experience as social expectations increase, may lead them to resort more frequently to destructive acts or displays of power in moments of anger and frustration; the conflict between seeking developmental independence and facing restrictive rules may also increase destructive tendencies. Oppositional defiant disorder is often seen in children of this age, and conflict may be especially common in families with high emotional expressivity, which may also lead to increased destructive behavior.⁴⁷

When CDSS subscale scores were examined by child sex, parents of male children reported higher impulsivity scores, whereas parents of female children reported higher archaic anxiety scores (Table 6). Archaic anxiety is associated with death, loss, and damage, whereas impulsivity is associated with hyperactivity and aggression.²⁷ Our results are consistent with tendencies toward excessive mobility and extroverted behavior in boys, and with the separation anxiety and fear-oriented reactions often associated with girls. Studies conducted in similar groups also showed that behaviors such as aggression, hyperactivity, and inattention were more common in boys, whereas fear and separation anxiety were more common in girls.^{25,26}

Study Limitations

Parents' reluctance and negative attitudes toward participating in surveys related to preschool depression constitute significant limitations for studies in this field. Furthermore, the scarcity of data on depressive symptoms in the 3- to 6-year age group posed a challenge when comparing our findings with the existing literature.

The present study has several methodological limitations. Because the research primarily relied on univariate and bivariate analyses, potential confounding variables could not be controlled, nor could independent predictors be identified. Consequently, the findings should be considered descriptive and hypothesis-generating, and future studies are recommended to employ multivariate analyses. Additionally, the lack of adjustments for multiple comparisons may have inflated the risk of type I errors. Although effect sizes were reported, some findings may reflect spurious associations; thus, the results must be interpreted with caution. The small sample size in certain subgroup analyses, particularly among children with chronic illnesses, also limits the reliability and generalizability of those specific findings.

Regarding psychometrics, internal consistency was low in certain CDSS subscales, particularly impaired cognitive functioning and fantasy-reality confusion indicating limitations in measuring these constructs. Furthermore, because the study was conducted within a single sociocultural region, its generalizability is limited, and the cross-sectional design precludes causal inference.

Nevertheless, this study contributes valuable data to the limited literature on depressive symptoms in preschool children. Using parent-reported data collected in the natural home environment, the study yields findings with high ecological validity. Ultimately, these results have the potential to inform and enhance early intervention and preventive mental health initiatives.

CONCLUSION

This study demonstrated that both parental psychosocial characteristics and children's individual and environmental characteristics significantly affected parent-reported depressive symptoms among children aged 3-6 years. Parents with chronic diseases reported higher depressive symptom subscale scores in their children. When compared across age and sex groups, fantasy-reality confusion scores were highest among 5-year-old children, aggression/destructiveness scores were highest among 6-year-old children, impulsivity scores were highest among boys, and archaic anxiety scores were highest among girls. Subscale scores for separation anxiety and somatization were higher for children with an allergic or other chronic disease who used regular medication. These findings suggest that mental health professionals and other health

care providers working with children should develop early diagnostic and intervention programs that take these risk factors into account. Providing information and developing psychoeducational programs for parents may also effectively support their children's mental health. Furthermore, initiatives to enhance social cohesion may be beneficial.

MAIN POINTS

- Parent-reported depressive symptoms in preschool children were associated with both child-related and parental psychosocial factors.
- Children with allergic or chronic diseases, and those with poor academic performance, showed higher levels of depressive symptoms.
- Parental chronic illness was associated with increased depressive symptom subscale scores in children.
- Boys showed higher impulsivity scores, whereas girls showed higher archaic anxiety scores.

ETHICS

Ethics Committee Approval: Approved the study protocol Ağrı İbrahim Çeçen University Scientific Research Ethics Committee (approval number: 406, date: 28.11.2024).

Informed Consent: All participants provided written informed consent.

Footnotes

Authorship Contributions

Surgical and Medical Practices: M.C., Y.A., B.D.G., Concept: M.C., Y.A., B.D.G., Design: M.C., Y.A., B.D.G., Data Collection and/or Processing: Y.A., Analysis and/or Interpretation: M.C., Literature Search: M.C., Y.A., B.D.G., Writing: M.C., Y.A., B.D.G.

DISCLOSURES

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

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

Declaration on the Use of Artificial Intelligence (AI): No artificial intelligence tools were used in the preparation of this manuscript.

REFERENCES

1. Baughman N, Prescott SL, Rooney R. The prevention of anxiety and depression in early childhood. *Front Psychol*. 2020; 11: 517896.
2. Gawlik KS, Melnyk BM, Tan A. Burnout and mental health in working parents: risk factors and practice implications. *J Pediatr Health Care*. 2025; 39(1): 41-50.
3. Kiliç C, Üstün B, Uluğ ÖŞ, Ak S, Ulusoy M, Kamilşli S, et al. Prevalence, persistence, and severity of 12-month and 30-day dsm-5 disorders in the Turkish World Mental Health Survey. *Turk Psikiyatri Derg*. 2026; 37: 180-91.
4. World Health Organization. *World Mental Health Report*. WHO: 2022. Access date: 28 February 2026. Available from: <https://www.who.int/teams/mental-health-and-substance-use/world-mental-health-report>
5. Domènech-Llaberia E, Araneda N, Xifró A, Riba MD. Escala de depresión preescolar para maestros (ESDM 3-6): primeros resultados de validación. *Psicothema*. 2003; 15(4): 575-80. Spanish.

6. Ramdhoney-Dowlot K, Balloo K, Morgül E, Essau CA. Prevalence of anxiety and depression among children and adolescents in low- and middle-income countries-a systematic review. *Psychiatr Res Clin Pract*. 2025; 7(4): 218-43.
7. Donohue MR, Whalen DJ, Gilbert KE, Hennefield L, Barch DM, Luby J. Preschool depression: a diagnostic reality. *Curr Psychiatry Rep*. 2019; 21(12): 128.
8. Whalen DJ, Sylvester CM, Luby JL. Depression and anxiety in preschoolers: a review of the past 7 years. *Child Adolesc Psychiatr Clin N Am*. 2017; 26(3): 503-22.
9. Juul EML, Hjemdal O, Aune T. Prevalence of depressive symptoms among older children and young adolescents: a longitudinal population-based study. *Scand J Child Adolesc Psychiatr Psychol*. 2021; 9: 64-72.
10. Lewis AJ, Sae-Koew JH, Toumbourou JW, Rowland B. Gender differences in trajectories of depressive symptoms across childhood and adolescence: a multi-group growth mixture model. *J Affect Disord*. 2020; 260: 463-72.
11. Kieling C, Buchweitz C, Caye A, Silvani J, Ameis SH, Brunoni AR, et al. Worldwide prevalence and disability from mental disorders across childhood and adolescence: evidence from the global burden of disease study. *JAMA Psychiatry*. 2024; 81(4): 347-56.
12. Hoyniak CP, Donohue MR, Hennefield L, Whalen DJ. Preschool mood disorders: a review of the literature from 2017 to 2024. *Child Adolesc Psychiatr Clin N Am*. 2025; 34(2): 325-37.
13. Silver J, Olino TM, Carlson GA, Dougherty LR, Bufferd SJ, Klein DN. Depression in 3/6-year-old children: clinical and psychosocial outcomes in later childhood and adolescence. *J Child Psychol Psychiatry*. 2022; 63(9): 984-91.
14. Koyuncu Z, Zabcı N, Seçen Yazıcı M, Sandıkçı T, Çetin Kara H, Doğangün B. Evaluating the association between developmental language disorder and depressive symptoms in preschool children. *Applied Neuropsychology: Child*. 2026; 15(2): 159-68.
15. Power J, Watson S, Chen W, Lewis AJ, van IJendoorn MH, Galbally M. Maternal emotional availability and perinatal depressive symptoms as predictors of early childhood executive function. *J Affect Disord*. 2024; 365: 332-40.
16. Camacho NL, Fowler CH, Gaffrey MS. Dimensions of depressive symptoms in young children: factor analysis of the preschool feelings checklist-scale. *Assessment*. 2025; 32(4): 544-60.
17. Genç M, Tolan Ö. Play therapy practices in psychological and developmental disorders that are common in preschool period. *Current Approaches in Psychiatry*. 2021; 13(2): 207-31.
18. Luby JL, Gaffrey MS, Tillman R, April LM, Belden AC. Trajectories of preschool disorders to full DSM depression at school age and early adolescence: continuity of preschool depression. *Am J Psychiatry*. 2014; 171(7): 768-76.
19. McGinnis EW, Anderau SP, Hruschak J, Gurchiek RD, Lopez-Duran NL, Fitzgerald K, et al. Giving voice to vulnerable children: machine learning analysis of speech detects anxiety and depression in early childhood. *IEEE J Biomed Health Inform*. 2019; 23(6): 2294-301.
20. Carneiro A, Soares I, Rescorla L, Dias P. Meta-analysis on parent-teacher agreement on preschoolers' emotional and behavioural problems. *Child Psychiatry Hum Dev*. 2021; 52(4): 609-18.
21. Egger HL, Angold A. Common emotional and behavioral disorders in preschool children: presentation, nosology, and epidemiology. *J Child Psychol Psychiatry*. 2006; 47(3-4): 313-37.
22. Wakschlag LS, Sherlock P, Blackwell CK, Burns JL, Krogh-Jespersen S, Gershon RC, et al. Modeling the normal:abnormal spectrum of early childhood internalizing behaviors: a clinical-developmental approach for the multidimensional assessment profiles internalizing dimensions. *Int J Methods Psychiatr Res*. 2023; 32(51): e1987.
23. World Health Organization, UNICEF. WHO and UNICEF launch guidance to improve access to mental health care for children and young people. 2024. Available from: <https://www.who.int/news/item/09-10-2024-who-and-unicef-launch-guidance-to-improve-access-to-mental-health-care-for-children-and-young-people>
24. Yazıcıoğlu Y, Erdoğan S. SPSS uygulamalı bilimsel araştırma yöntemleri. Ankara: Detay Yayıncılık; 2004.
25. Derdiyok ZS, Kahraman ÖG. Investigation of preschool children's anxiety in terms of parent emotion socialization behaviors. *Journal of Early Childhood Studies*. 2024; 8(1): 100-29.
26. Işık E. Examining behavioral problems in early childhood in terms of various variables. *Uludağ Üni Fen-Edeb Fak Sos Bil Derg*. 2021; 22(40): 183-226.
27. Erol E, Zabcı N, Şimşek ÖF. Development of child depressive symptoms assessment scale. *Anadolu Psikiyatri Derg*. 2020; 21(Suppl 2): 14-20.
28. Onat B, Özkan E, Algedik P. The independent protective role of paternal involvement in the association between maternal containment and depressive symptoms in children aged 3-6. *J Med Palliat Care*. 2026; 7(4): 710-6.
29. Kaya F, Aktaş S. The contribution of recreation areas developed in the city of agri and its close to urban livability. *Ağrı İbrahim Çeçen Üni Sos Bil Ens Derg*. 2023; 9(2): 271-300.
30. Misirli A, Fotakopoulou O, Dardanou M, Komis V. The impact of touchscreen digital exposure on children's social development and communication: a systematic review. *Front Psychol*. 2025; 16: 1613625.
31. Zabcı N, Erol E. Relationship between depressive symptoms of children 3-7 years and the containing functions of their moms. *AYNA Clin Psychol J*. 2021; 8(3): 510-34.
32. Elliott L, Thompson KA, Fobian AD. A systematic review of somatic symptoms in children with a chronically ill family member. *Psychosom Med*. 2020; 82(4): 366-76.
33. Razaz N, Joseph KS, Boyce WT, Guhn M, Forer B, Carruthers R, et al. Children of chronically ill parents: relationship between parental multiple sclerosis and childhood developmental health. *Mult Scler*. 2016; 22(11): 1452-62.
34. Zhang Y, Jiang M. Parent-child mismatch in educational expectations and depressive symptoms among Chinese adolescents. *Healthcare*. 2024; 12(17): 1792.
35. Zhao H, Zhu X, Li W, Lin X. When parents worry: how parental educational anxiety impacts adolescent academic success through depression, self-efficacy, and social media. *Int J Ment Health Promot*. 2025; 27(4): 517-40.
36. Mundy LK, Canterford L, Moreno-Betancur M, Hoq M, Viner RM, Bayer JK, et al. Learning outcomes in primary school children with emotional problems: a prospective cohort study. *Child Adolesc Ment Health*. 2023; 28(3): 377-84.
37. Gao Y, Hu D, Peng E, Abbey C, Ma Y, Wu CI, et al. Depressive symptoms and the link with academic performance among rural Taiwanese children. *Int J Environ Res Public Health*. 2020; 17(8): 2778.
38. Massetti GM, Lahey BB, Pelham WE, Loney J, Ehrhardt A, Lee SS, Kipp H. Academic achievement over 8 years among children who met modified criteria for attention-deficit/hyperactivity disorder at 4-6 years of age. *J Abnorm Child Psychol*. 2008; 36: 399-410.
39. Battaglia M, Garon-Carrier G, Côté SM, Dionne G, Touchette E, Vitaro F, et al. Early childhood trajectories of separation anxiety: bearing on mental health, academic achievement, and physical health from mid-childhood to preadolescence. *Depress Anxiety*. 2017; 34(10): 918-27.
40. Edvinsson Sollander S, Fabian H, Sarkadi A, Salari R, Fält E, Dahlberg A, et al. Asthma and allergies correlate with mental health problems in preschool children. *Acta Paediatr*. 2021; 110(5): 1601-9.
41. Kern C, Wan J, LeWinn KZ, Ramirez FD, Lee Y, McCulloch CE, et al. Association of atopic dermatitis and mental health outcomes across childhood: a longitudinal cohort study. *JAMA Dermatol*. 2021; 157(10): 1200-8.

42. Pinquart M, Shen Y. Depressive symptoms in children and adolescents with chronic physical illness: an updated meta-analysis. *J Pediatr Psychol.* 2011; 36(4): 375-84.
43. Barker MM, Beresford B, Fraser LK. Incidence of anxiety and depression in children and young people with life-limiting conditions. *Pediatr Res.* 2023; 93(7): 2081-90.
44. Boris P, Kovács KE, Nagy BE. The comparative study of chronically ill and healthy children and adolescents in the light of their general mental health. *Sci Rep.* 2024; 14(1): 6754.
45. Jansen M, Bodden DHM, Muris P, van Doorn M, Granic I. Measuring anxiety in children: the importance of separate mother and father reports. *Child Youth Care Forum.* 2017; 46(5): 643-59.
46. Woolley JD, Ghossainy ME. Revisiting the fantasy–reality distinction: children as naïve skeptics. *Child Development.* 2013; 84(5): 1496-510.
47. Demir Gökmen B, Yılmaz G. Çocuklarda yıkıcı dürtü denetimli kontrol ve davranım bozuklukları ve hemşirelik yaklaşımı. In: Kardaş Özdemir F, Günaydın N, Küçük Alemdar D, editors. *Çocuklarda psikososyal sorunlar ve hemşirelik yaklaşımı.* Ankara: Akademisyen Kitabevi; 2025. p. 40-73.

The Impact of Laparoscopic and Transvaginal Vaginal Cuff Closure Techniques on Surgical and Functional Outcomes in Total Laparoscopic Hysterectomy

✉ Mehmet Emre Peker¹, ✉ Ufuk Atlihan², ✉ Can Ata², ✉ Duygu Uçar Kartal¹, ✉ Haydar Kaya¹

¹Clinic of Obstetrics and Gynecology, Merkezefendi State Hospital, Manisa, Türkiye

²Clinic of Obstetrics and Gynecology, Buca Seyfi Demirsoy Training and Research Hospital, İzmir, Türkiye

Abstract

BACKGROUND/AIMS: To evaluate whether the technique used for vaginal cuff closure after total laparoscopic hysterectomy (TLH) influences perioperative outcomes and early postoperative sexual function.

MATERIALS AND METHODS: Clinical records from women undergoing TLH for benign gynecological indications at Buca Seyfi Demirsoy Training and Research Hospital during the period from January 2021 to December 2024 were retrospectively evaluated. Patients were assigned to either a transvaginal suturing group (n=71) or a laparoscopic intracorporeal suturing group (n=55), according to the route of vaginal cuff closure. Demographic characteristics, operative variables, perioperative complications, vaginal length measurements, and female sexual function index (FSFI) scores were analyzed. Sexual function was assessed before surgery and again three months postoperatively.

RESULTS: No statistically significant differences were observed between the groups regarding baseline characteristics. Women in the laparoscopic closure group demonstrated significantly greater postoperative vaginal length compared with those who underwent transvaginal closure. The duration of surgery and the length of hospitalization were similar. Although both techniques were associated with postoperative improvement in FSFI scores, the degree of improvement was significantly higher among patients who received laparoscopic cuff closure.

CONCLUSION: Intracorporeal laparoscopic closure of the vaginal cuff following TLH is associated with improved preservation of vaginal length and greater enhancement of early postoperative sexual function compared with the transvaginal approach.

Keywords: Total laparoscopic hysterectomy, laparoscopic cuff closure, vaginal cuff closure, sexual function, FSFI

INTRODUCTION

Hysterectomy continues to represent a major component of surgical management for benign gynecologic disorders.¹ With ongoing improvements in endoscopic instrumentation and surgical training, total laparoscopic hysterectomy (TLH) has become increasingly prevalent over the last decade.² Compared with open abdominal hysterectomy, minimally invasive approaches are associated with

reduced postoperative discomfort, shorter inpatient stay, earlier mobilization, and improved cosmetic satisfaction.³ Nevertheless, TLH remains a complex operation that demands advanced laparoscopic proficiency and is characterized by a substantial learning curve.⁴

Reconstruction of the vaginal vault following uterine removal constitutes a pivotal stage of TLH because it directly influences both perioperative safety and long-term outcomes.⁵ Suboptimal

To cite this article: Peker ME, Atlihan U, Ata C, Uçar Kartal D, Kaya H. The impact of laparoscopic and transvaginal vaginal cuff closure techniques on surgical and functional outcomes in total laparoscopic hysterectomy. Cyprus J Med Sci. 2026;11(5):353-358

ORCID IDs of the authors: M.E.P. 0000-0002-3344-2935; U.A. 0000-0002-2109-1373; C.A. 0000-0002-0841-0480; D.U.K. 0000-0003-4751-5944; H.K. 0000-0001-9021-5125.



Corresponding author: Mehmet Emre Peker

E-mail: epekerrr@gmail.com

ORCID ID: orcid.org/0000-0002-3344-2935

Received: 01.01.2026

Accepted: 10.07.2026

Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.

This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

approximation of cuff tissues may predispose patients to hemorrhage, infectious complications, granulation tissue formation, or separation of the vaginal vault edges.⁶ Furthermore, impaired healing at the cuff site may compromise vaginal dimensions and apical support, thereby affecting postoperative sexual health.⁷

Currently, two principal strategies are employed for vaginal cuff reconstruction during TLH: intracorporeal laparoscopic suturing and transvaginal suturing after specimen extraction.⁸ Laparoscopic suturing allows surgeons to realign tissues under magnified visualization, facilitating circumferential closure and direct hemostatic control.⁹ However, this approach may be technically demanding, particularly during the initial learning phase, and has been reported to extend operative duration.¹⁰

Conversely, closure through the vaginal route is often considered technically straightforward and may shorten the procedure, particularly for surgeons with limited laparoscopic suturing experience.¹¹ Despite this, the restricted operative field inherent to the vaginal approach can hinder optimal hemostasis. Prior investigations have suggested that transvaginal closure may be associated with increased rates of early bleeding or infectious sequelae, whereas laparoscopic closure may promote more uniform tissue healing.^{12,13}

The integrity of the vaginal cuff is closely linked to postoperative sexual well-being. Maintenance of vaginal length, elasticity, and structural support is known to influence parameters such as dyspareunia, orgasmic function, and overall sexual satisfaction.¹⁴ Accordingly, minimizing tissue trauma during cuff reconstruction is regarded as a key determinant of postoperative quality of life.¹⁵

Although multiple studies have examined the impact of cuff closure technique on surgical outcomes, their conclusions remain inconsistent. While some authors report reduced complication rates with laparoscopic suturing despite longer operative times, others favor the practicality of the transvaginal approach, albeit with concerns regarding postoperative inflammatory changes.^{1,8} Moreover, several reports indicate superior preservation of vaginal length and improved sexual function following laparoscopic closure.^{9,14}

Given the ongoing debate surrounding the optimal method of vaginal cuff closure in TLH, this retrospective study was designed to compare laparoscopic and transvaginal techniques in terms of operative parameters, perioperative morbidity, and postoperative sexual function. We aimed to contribute additional evidence to inform surgical decision-making for TLH.

MATERIALS AND METHODS

This retrospective observational study included women who underwent TLH for non-malignant gynecologic conditions at Buca Seyfi Demirsoy Training and Research Hospital between January 2021 and December 2024. Ethical approval for this study was obtained from the Buca Seyfi Demirsoy Training and Research Hospital Non-Interventional Research Ethics Committee (approval no: 2025/526, date: 26.11.2025). The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki.

A total of 126 consecutive cases were eligible for analysis. Patients with malignant pathology, those who required additional pelvic reconstructive procedures, those converted to open surgery, or those with missing perioperative information were excluded. Participants

were categorized based on the vaginal cuff closure method applied during surgery: transvaginal suturing (Group 1, n=71) and intracorporeal laparoscopic suturing (Group 2, n=55).

To minimize inter-operator variability, all surgeries were performed by the same experienced team using a consistent minimally invasive surgical technique. After laparoscopic coagulation of the uterine arteries and completion of colpotomy, the uterus was extracted through the vaginal canal. In Group 1, the vaginal vault was approximated via the vaginal route with absorbable sutures using either an interrupted or a running technique. In Group 2, closure was completed endoscopically under laparoscopic guidance, most often employing barbed suture material.

Data collected from medical records included demographic characteristics, obstetric history, body mass index (BMI), surgical indications, prior pelvic operations, operative duration, perioperative complications, length of hospitalization, and vaginal length measurements obtained before and after surgery. Sexual function was evaluated using the female sexual function index (FSFI) questionnaire, which was administered preoperatively and again three months postoperatively. Although the present analysis was limited to early postoperative outcomes at three months, extended follow-up assessments at 6 and 12 months are planned to further evaluate the long-term impact of vaginal cuff closure techniques on sexual function and anatomical outcomes. This validated instrument contains 19 items covering six functional domains: desire, arousal, lubrication, orgasm, satisfaction, and pain, yielding a total score between 2 and 36.

Statistical Analysis

Data analysis was performed using IBM SPSS Statistics for Windows, version 29.0 (IBM Corp., Armonk, NY, USA). Descriptive statistics are reported as mean $\pm$ standard deviation (SD) for continuous variables and as numbers with percentages for categorical variables. The normality of continuous variables was determined by combining the Shapiro-Wilk test with graphical assessment using histograms and Q-Q plots. Equality of variances between groups was verified with Levene's test before inferential analyses. When the assumptions for parametric testing were met, between-group comparisons were performed using the appropriate parametric methods. Differences in categorical variables were evaluated using either the chi-square test or Fisher's exact test, depending on the expected cell counts. Statistical significance was defined as a two-sided p-value <0.05 .

Within-group changes in FSFI scores from baseline to the third postoperative month were examined, and the magnitudes of these changes were then compared between the two study groups. Statistical significance was defined as a two-sided probability threshold of 0.05. As this investigation was retrospective, no prospective sample size estimation was undertaken before data collection. Following the completion of the analyses, a post-hoc power calculation was performed using the change in FSFI score as the primary endpoint. Based on an observed mean difference of 3.28, a pooled SD of 8.4, group sizes of 71 and 55 participants, and a two-sided significance level of 0.05, the estimated statistical power was approximately 80%. These findings indicate that the available sample size was adequate to detect clinically relevant differences between the study groups.

RESULTS

A total of 126 patients were included in the study; vaginal cuff closure was performed transvaginally in 71 patients and laparoscopically in 55 patients. No statistically significant differences were identified between the two groups with respect to age, BMI, gravidity, or parity (all $p>0.05$). The mean age was 45.8 ± 10.4 years in the transvaginal suture group and 47.1 ± 9.3 years in the laparoscopic suture group ($p=0.42$). The mean BMI values were 28.1 ± 4.1 and 27.4 ± 3.8 , respectively ($p=0.35$) (Table 1).

The distribution of hysterectomy indications was similar between the two groups, with no statistically significant differences detected. The most common indication in both groups was myoma uteri (transvaginal: 40.8%; laparoscopic: 38.2%; $p=0.76$), followed by adenomyosis (35.2% and 36.4%, respectively) and adnexal masses (14.1% and 16.4%, respectively). The rates of simple or complex endometrial hyperplasia were also similar between the groups ($p=0.88$). The proportion of patients with prior pelvic surgery was high in both groups (transvaginal 66.2%; laparoscopic 67.3%), and no statistically significant difference was observed ($p=0.89$) (Table 2).

Preoperative vaginal length values were similar between the groups (transvaginal: 7.98 ± 0.97 cm; laparoscopic: 8.02 ± 1.02 cm; $p=0.79$). However, postoperative vaginal length measurements were significantly higher in the laparoscopic group (7.52 ± 0.96 cm vs. 7.18 ± 0.83 cm,

$p=0.04$). Mean operative time and duration of hospitalization were comparable between the study groups, with no statistically significant differences observed ($p=0.29$ and $p=0.18$, respectively) (Table 3).

No significant between-group differences were identified in the incidence of either intraoperative or postoperative complications. In the transvaginal group, hematoma occurred in 4 patients (5.6%) and bladder injury occurred in 1 patient (1.4%), whereas in the laparoscopic group hematoma and bladder injury were observed in 2 (3.6%) and 1 (1.8%) patients, respectively ($p>0.05$). The rates of cuff infection, urinary tract infection, and trocar site infection were low, with no significant differences between the groups (all $p>0.05$) (Table 4).

Postoperatively, FSFI scores increased in both groups. In the transvaginal suture group, the mean FSFI score rose from 22.54 ± 9.35 preoperatively to 24.76 ± 8.54 at the third postoperative month, while in the laparoscopic suture group, it increased from 21.41 ± 8.76 to 26.91 ± 8.13 . The mean increase in FSFI score was calculated as 2.22 in the transvaginal group and 5.50 in the laparoscopic group. The difference between the two groups was statistically significant ($p=0.045$) (Tables 5 and 6). The observed improvement in FSFI scores corresponded to a moderate effect size (Cohen's $d=0.42$; 95% confidence interval, 0.05-0.79). This magnitude of effect suggests that laparoscopic vaginal cuff closure was associated with a clinically relevant advantage in postoperative sexual function.

Table 1. Demographic and clinical characteristics of the groups

Parameters	Vaginal cuff closure performed transvaginally (n=71)	Vaginal cuff closure performed laparoscopically (n=55)	p-value
Age	45.8±10.4	47.1±9.3	0.42
BMI	28.1±4.1	27.4±3.8	0.35
Gravida	3.5±1.6	3.2±1.3	0.24
Parity	2.4±1.3	2.3±1.1	0.58

Since all variables were continuous, the Independent Samples t-test was applied.

BMI: Body mass index.

Table 2. Indications for hysterectomy and history of pelvic surgery in both groups

Indications for hysterectomy	Vaginal cuff closure performed transvaginally (n=71)	Vaginal cuff closure performed laparoscopically (n=55)	p-value
Myoma	29 (40.8%)	21 (38.2%)	0.76
Adenomyosis	25 (35.2%)	20 (36.4%)	0.89
Adnexal mass	10 (14.1%)	9 (16.4%)	0.73
Simple or complex hyperplasia	7 (9.9%)	5 (9.1%)	0.88

History of pelvic surgery

Yes	47 (66.2%)	37 (67.3%)	0.89
No	24 (33.8%)	18 (32.7%)	

Since this table includes categorical variables (indications and presence or absence of surgical history), the chi-square (χ^2) test or Fisher's exact test (for small cell frequencies) was used.

Table 3. Comparison of vaginal length, operative time, and length of hospital stay between the two groups

Parameters	Vaginal cuff closure performed transvaginally (n=71)	Vaginal cuff closure performed laparoscopically (n=55)	p-value
Preoperative vaginal length (cm)	7.98±0.97	8.02±1.02	0.79
Postoperative vaginal length (cm)	7.18±0.83	7.52±0.96	0.04
Operative time (min)	60.8±14.94	57.8±15.16	0.29
Length of hospital stay (d)	2.4±1.14	2.67±1.21	0.18

Since this table includes continuous variables, the Independent Samples t-test was used.

The comparison of changes in FSFI domain scores between transvaginal and laparoscopic vaginal cuff closure techniques following TLH. Overall, improvements were observed in all domains in both groups, with consistently greater increases in the laparoscopic closure group (Table 6).

Statistically significant differences between groups were identified in the arousal ($p=0.047$), lubrication ($p=0.039$), and pain ($p=0.028$) domains, suggesting that laparoscopic closure may confer superior benefits for sexual arousal, vaginal lubrication, and reduction of dyspareunia. In contrast, the desire ($p=0.084$), orgasm ($p=0.052$), and satisfaction ($p=0.061$) domains did not reach statistical significance, although all showed a trend favoring the laparoscopic approach.

These results may indicate that laparoscopic vaginal cuff closure has a more pronounced positive impact on postoperative sexual function, particularly in domains closely related to physical comfort and physiological responses. The lack of significance in some domains could be attributed to sample size limitations or inter-individual variability in sexual function.

DISCUSSION

The present study demonstrated that, following TLH, intracorporeal laparoscopic vaginal cuff closure was associated with more favorable

early postoperative functional outcomes than transvaginal closure. Although postoperative FSFI scores improved significantly in both groups, the magnitude of improvement was greater among patients who underwent laparoscopic vaginal cuff closure. Furthermore, postoperative vaginal length was better preserved in this group, suggesting that the laparoscopic approach may offer additional anatomical and functional advantages.

Previous studies have demonstrated that postoperative sexual function following hysterectomy is influenced by several factors, including the surgical indication, operative technique, and the quality of vaginal cuff healing. Several reports have shown significant improvements in FSFI scores after TLH,¹⁶ while others have highlighted improvements particularly in the domains of sexual desire and orgasm, emphasizing the importance of preserving vaginal length for optimal postoperative sexual function.¹⁷ These findings are consistent with the results of the present study.

Intracorporeal suturing allows precise circumferential approximation of the vaginal tissues under magnified laparoscopic visualization, thereby facilitating controlled hemostasis and optimal tissue alignment. In a multicenter analysis, Uccella et al.¹⁸ demonstrated that laparoscopic vaginal cuff closure was associated with a lower incidence of vaginal cuff dehiscence compared with the transvaginal approach.

Table 4. Comparison of intraoperative and postoperative complications between the two groups

Intraoperative and postoperative complications	Vaginal cuff closure performed transvaginally (n=71)	Vaginal cuff closure performed laparoscopically (n=55)	p-value
Hematoma	4 (5.6%)	2 (3.6%)	0.42
Bladder injury	1 (1.4%)	1 (1.8%)	0.88
Cuff infection	2 (2.8%)	1 (1.8%)	0.72
Urinary tract infection	2 (2.8%)	1 (1.8%)	0.72
Trocar site infection	1 (1.4%)	1 (1.8%)	0.88

Since this table includes categorical variables (n, %), the chi-square (χ^2) test or, when necessary, Fisher's exact test was used for evaluation.

Table 5. Comparison of preoperative and postoperative FSFI scores between the two groups

	Vaginal cuff closure performed transvaginally (n=71)	Vaginal cuff closure performed laparoscopically (n=55)	p-value
Preoperative FSFI	22.54±9.35	21.41±8.76	
Postoperative 3 rd -month FSFI	24.76±8.54	26.91±8.13	
FSFI difference	2.22	5.5	0.045

Since these data are continuous numerical variables (FSFI scores), the Independent Samples t-test was used. Between-group comparisons of FSFI change scores were performed using the Independent Samples t-test.

FSFI: Female sexual function index.

Table 6. Comparison of changes in FSFI domain scores between transvaginal and laparoscopic vaginal cuff closure techniques

FSFI domain	Transvaginal closure (n=71) Δ	Laparoscopic closure (n=55) Δ	p-value
Desire	+0.29	+0.62	0.084
Arousal	+0.37	+0.81	0.047
Lubrication	+0.36	+0.94	0.039
Orgasm	+0.36	+0.91	0.052
Satisfaction	+0.36	+0.85	0.061
Pain	+0.48	+1.37	0.028
Total FSFI	+2.22	+5.50	0.045

Δ: Difference between postoperative and preoperative FSFI domain scores (3-month postoperative-baseline).

FSFI: Female sexual function index.

Similarly, another study reported a reduced rate of granulation tissue formation following laparoscopic suturing.¹⁹ These technical advantages may explain the superior preservation of vaginal length observed in the laparoscopic group.

Regarding sexual function, previous studies have reported that minimally invasive hysterectomy techniques are associated with better preservation of vaginal elasticity, which may contribute to improved postoperative sexual function.^{20,21} Consistent with these findings, our study demonstrated a significantly greater improvement in postoperative sexual function in women who underwent laparoscopic vaginal cuff closure.

Although the transvaginal approach has been described as technically less demanding and potentially associated with shorter operative times, particularly for less experienced surgeons,^{22,23} no significant difference in operative duration was observed between the two groups in our study. This finding is likely attributable to the standardized surgical technique and the experience of the operating surgeons.

Accurate and symmetrical reconstruction of the vaginal cuff is important not only for postoperative sexual function but also for long-term apical support. Previous studies have demonstrated superior apical support following laparoscopic vaginal cuff closure,²⁴ while prospective investigations have reported greater postoperative vaginal length and higher FSFI scores in patients undergoing this technique.²⁵ Our findings are consistent with these reports and further support the growing body of evidence suggesting that intracorporeal laparoscopic vaginal cuff closure provides both anatomical and functional advantages following TLH.

Study Limitations

The findings of the present study should be interpreted in light of several methodological limitations. The retrospective nature of the analysis precludes full control over potential confounding variables and introduces the possibility that patient selection or clinical decision-making influenced the observed outcomes. In addition, sexual function was assessed only during the early postoperative period, preventing the evaluation of longer-term functional trajectories.

Although all procedures were conducted within a single surgical unit following uniform protocols, individual technical nuances and the influence of cumulative surgical experience may still have contributed to outcome variability. Moreover, the limited sample size reduces the ability to reliably identify infrequent adverse events, such as vaginal cuff separation or serious postoperative infections. Finally, because patients were not randomly allocated to closure techniques, definitive causal inferences regarding the superiority of one method over the other cannot be made. Furthermore, statistical methods designed for longitudinal data, including repeated-measures analysis of variance and mixed-effects modeling, would have enabled a more comprehensive evaluation of temporal changes, group differences, and their interaction. However, the retrospective nature of the study, together with the available sample size, precluded the use of these approaches. Accordingly, the reported effect size estimates should be interpreted with appropriate caution.

CONCLUSION

This study demonstrates that intracorporeal laparoscopic reconstruction of the vaginal cuff after TLH is associated with better preservation of vaginal anatomy and greater improvement in postoperative sexual function than transvaginal suturing. In settings where sufficient laparoscopic expertise is available, this approach represents a reliable and clinically advantageous option that may contribute to improved recovery and overall patient-reported outcomes.

MAIN POINTS

- Intracorporeal laparoscopic vaginal cuff closure preserved postoperative vaginal length more effectively than transvaginal closure.
- Both closure techniques improved postoperative sexual function; however, laparoscopic closure resulted in significantly greater improvement in total female sexual function index scores.
- Perioperative complication rates, operative time, and length of hospital stay were comparable between the two techniques.
- Laparoscopic vaginal cuff closure may provide superior early functional outcomes without increasing surgical morbidity.
- When adequate laparoscopic expertise is available, intracorporeal cuff closure represents a valuable option for optimizing patient-reported postoperative outcomes.

ETHICS

Ethics Committee Approval: Ethical approval for this study was obtained from the institutional Buca Seyfi Demirsoy Training and Research Hospital Non-Interventional Research Ethics Committee (approval no: 2025/526, date: 26.11.2025).

Informed Consent: This is a retrospective observational study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: M.E.P., U.A., Concept: M.E.P., C.A., Design: M.E.P., U.A., Data Collection and/or Processing: C.A., H.K., Analysis and/or Interpretation: U.A., D.U.K., H.K., Literature Search: M.E.P., D.U.K., H.K., Writing: M.E.P., C.A.

DISCLOSURES

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

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

Declaration of Generative AI and AI-assisted Technologies in the Writing Process: During the preparation of this work, the author(s) used ChatGPT to improve “readability” and “language”. After using this tool/service, the author(s) reviewed and edited the content as needed and took full responsibility for the content of the publication. Artificial intelligence has been used to assist with “text editing”.

REFERENCES

1. Bastu E, Yasa C, Dural O, Ozgor BY, Yilmaz G, Gungor Ugurlucan F, et al. Comparison of 2 methods of vaginal cuff closure at laparoscopic hysterectomy and their effect on female sexual function and vaginal length: a randomized clinical study. *J Minim Invasive Gynecol*. 2016; 23(6): 986-93.
2. Uçar MG, İlhan TT, Şanlıkan F, Çelik Ç. Sexual functioning before and after vaginal hysterectomy to treat pelvic organ prolapse and the effects of vaginal cuff closure techniques: a prospective randomised study. *Eur J Obstet Gynecol Reprod Biol*. 2016; 206: 1-5.
3. Uccella S, Capozzi VA, Ricco' M, Perrone E, Zanello M, Ferrari S, et al. Sexual function following Laparoscopic versus transvaginal closure of the vaginal vault after laparoscopic hysterectomy: secondary analysis of a randomized trial by the Italian Society of Gynecological Endoscopy using a validated questionnaire. *J Minim Invasive Gynecol*. 2020; 27(1): 186-94.
4. Pacis MM, Harkins G, Stetter C, Kunselman A, Benton A, Deimling T. A randomized controlled trial comparing horizontal vs vertical cuff closure during laparoscopic hysterectomy: impacts on total vaginal length and sexual function. *J Minim Invasive Gynecol*. 2017; 24(Suppl 7): S146.
5. Kalkan Ü, Bakay K. A multimodal concept for vaginal cuff closure by modification of the Bakay technique in total laparoscopic hysterectomy: a randomized clinical study. *BMC Womens Health*. 2022; 22(1): 6.
6. Gracia M, Martínez-Zamora MA, Castelo-Branco C, Carmona F. The influence of laparoscopic benign hysterectomy in sexual function. *Clinical and Experimental Obstetrics & Gynecology*. 2023; (50)2: 38.
7. Lonnée-Hoffmann R, Pinas I. Effects of Hysterectomy on sexual function. *Curr Sex Health Rep*. 2014; 6(4): 244-51.
8. Förster CE, Calabretti I, Gubser L, Schötzau A, Fellmann-Fischer B, Heinzelmann-Schwarz V, et al. Comparison of different suture techniques for laparoscopic vaginal cuff closure. *Sci Rep*. 2024; 14(1): 4860.
9. Elmetwaly SMM, Farouk Midan M, Megahed AEM. Laparoscopic versus traditional vaginal cuff closure at laparoscopic hysterectomy and their effect on female sexual function and vaginal length: a randomized clinical study. *IJMA*. 2026; 8(2): 6632-9.
10. Smith K, Caceres A. Vaginal cuff closure in minimally invasive hysterectomy: a review of training, techniques, and materials. *Cureus*. 2017; 9(10): e1766.
11. Hill AM, Davis KM, Clark-Donat L, Hammons LM, Azodi M, Silasi DA. The effect of vertical versus horizontal vaginal cuff closure on vaginal length after laparoscopic hysterectomy. *J Minim Invasive Gynecol*. 2017; 24(1): 108-13.
12. Cronin B, Sung VW, Matteson KA. Vaginal cuff dehiscence: risk factors and management. *Am J Obstet Gynecol*. 2012; 206(4): 284-8.
13. Ross WT, Meister MR, Shepherd JP, Olsen MA, Lowder JL. Utilization of apical vaginal support procedures at time of inpatient hysterectomy performed for benign conditions: a national estimate. *Am J Obstet Gynecol*. 2017; 217(4): 436.e1-436.e8.
14. Nezhat C, Kennedy Burns M, Wood M, Nezhat C, Nezhat A, Nezhat F. Vaginal cuff dehiscence and evisceration: a review. *Obstet Gynecol*. 2018; 132(4): 972-85.
15. Schimpf MO, Harvie HS, Omotosho TB, Epstein LB, Jean-Michel M, Olivera CK, et al; Society of Gynecologic Surgeons Fellows' Pelvic Research Network. Does vaginal size impact sexual activity and function? *Int Urogynecol J*. 2010; 21(4): 447-52.
16. El-Toukhy TA, Hefni M, Davies A, Mahadevan S. The effect of different types of hysterectomy on urinary and sexual functions: a prospective study. *J Obstet Gynaecol*. 2004; 24(4): 420-5.
17. Beyan E, İnan AH, Emirdar V, Budak A, Tutar SO, Kanmaz AG. Comparison of the effects of total laparoscopic hysterectomy and total abdominal hysterectomy on sexual function and quality of life. *Biomed Res Int*. 2020; 2020: 8247207.
18. Uccella S, Cromi A, Bogani G, Ghezzi F. Vaginal cuff dehiscence after laparoscopic hysterectomy. *Eur J Obstet Gynecol Reprod Biol*. 2012; 161(1): 113; author reply 113-4.
19. Uccella S, Ghezzi F, Mariani A, Cromi A, Bogani G, Serati M, Bolis P. Vaginal cuff closure after minimally invasive hysterectomy: our experience and systematic review of the literature. *Am J Obstet Gynecol*. 2011; 205(2): 119.e1-12.
20. Danesh M, Hamzehgardeshi Z, Moosazadeh M, Shabani-Asrami F. The effect of hysterectomy on women's sexual function: a narrative review. *Med Arch*. 2015; 69(6): 387-92.
21. Radosa JC, Meyberg-Solomayer G, Kastl C, Radosa CG, Mavrova R, Gräber S, et al. Influences of different hysterectomy techniques on patients' postoperative sexual function and quality of life. *J Sex Med*. 2014; 11(9): 2342-50.
22. Agarwal M, Sinha S, Singh S, Haripriya H, Simran S. Vaginal vault closure following total laparoscopic hysterectomy: laparoscopic versus conventional technique - a comparative study. *Gynecol Minim Invasive Ther*. 2023; 13(2): 90-4.
23. Köhler C, Hertel H, Herrmann J, Marnitz S, Mallmann P, Favero G, et al. Laparoscopic radical hysterectomy with transvaginal closure of vaginal cuff - a multicenter analysis. *Int J Gynecol Cancer*. 2019; 29(5): 845-50.
24. Hwang JH, Lee JK, Lee NW, Lee KW. Vaginal cuff closure: a comparison between the vaginal route and laparoscopic suture in patients undergoing total laparoscopic hysterectomy. *Gynecol Obstet Invest*. 2011; 71(3): 163-9.
25. Dundar B, Dincgez Cakmak B, Aykanat Y, Macunluoglu AC. Laparoscopic versus vaginal cuff closure in laparoscopic hysterectomy: does it affect female sexuality? *Ginekol Pol*. 2019; 90(7): 365-70.

Polycystic Ovary Syndrome Detection Using Gradient Boosting Algorithms and Explainable Artificial Intelligence: A Comparative Study

Ümran Karabulut Doğan¹, Şeyma Yaşar², Erhan Hüseyin Cömert³, Erol Karakaş⁴, Ozan Doğan⁵

¹Clinic of Obstetrics and Gynecology, Acıbadem Kayseri Hospital, Kayseri, Türkiye

²Department of Biostatistics and Medical Informatics, İnönü University Faculty of Medicine, Malatya, Türkiye

³Clinic of Obstetrics and Gynecology, Special Women's Health Clinic, İstanbul, Türkiye

⁴Clinic of Obstetrics and Gynecology, Special Women's Health Clinic, Kayseri, Türkiye

⁵Department of Obstetrics and Gynecology, İstanbul Nişantaşı University Faculty of Medicine, İstanbul, Türkiye

Abstract

BACKGROUND/AIMS: Polycystic ovary syndrome (PCOS) is one of the most common endocrine disorders affecting women of reproductive age, with a global prevalence ranging from 6% to 21%. Early and accurate diagnosis is critical to preventing long-term metabolic, reproductive, and cardiovascular complications. This study aims to develop and compare two gradient-boosting machine-learning, extreme gradient boosting (XGBoost) and light gradient boosting machine (LightGBM), for PCOS classification using clinical and hormonal features and to enhance model interpretability through explainable artificial intelligence techniques.

MATERIALS AND METHODS: A publicly available Kaggle dataset comprising 541 patients and 41 clinical, biochemical, and lifestyle features was utilised. After preprocessing-including median imputation and synthetic minority oversampling technique (SMOTE)-two ensemble classifiers were trained and validated using 5-fold stratified cross-validation, with SMOTE applied strictly within each training fold to prevent data leakage. Three baseline classifiers (logistic regression, random forest, and support vector machine) were included for comparison. SHapley Additive exPlanations (SHAP) summary plots, bar charts, waterfall diagrams, and local interpretable model-agnostic explanations (LIME) were employed for global and local model interpretability.

RESULTS: XGBoost achieved an accuracy of 91.74%, an area under the curve-receiver operating characteristic (AUC-ROC) of 0.9522 [95% confidence interval (CI): 0.910-0.992], and an F1-score of 0.8696, whereas LightGBM attained an accuracy of 89.91%, an AUC-ROC of 0.9551 (95% CI: 0.914-0.993), and an F1-score of 0.8451. Corrected 5-fold CV AUC scores were 0.9701 ± 0.0148 and 0.9682 ± 0.0161 , respectively. Both gradient boosting models outperformed all three baseline comparators. Calibration analysis yielded Brier scores of 0.071 and 0.079 for XGBoost and LightGBM, respectively, and decision curve analysis demonstrated positive net benefit over treat-all and treat-none strategies across clinically plausible threshold ranges. A sensitivity analysis excluding direct Rotterdam-criteria variables (follicle counts and cycle irregularity) yielded AUC-ROC values of 0.88 and 0.87, confirming residual discriminative capability of non-criterion features. SHAP analysis consistently identified right ovarian follicle number (Follicle No. R), hair growth (hirsutism), left ovarian follicle number (Follicle No. L), weight gain, and cycle irregularity as the five most influential predictors across both models.

To cite this article: Karabulut Doğan Ü, Yaşar Ş, Cömert EH, Karakaş E, Doğan O. Polycystic ovary syndrome detection using gradient boosting algorithms and explainable artificial intelligence: a comparative study. Cyprus J Med Sci.2026;11(5):359-368

ORCID IDs of the authors: Ü.K.D. 0000-0002-1935-3696; Ş.Y. 0000-0003-1300-3393; E.H.C. 0000-0003-1431-2294; E.K. 0000-0001-6050-6765; O.D. 0000-0002-0016-8749.



Corresponding author: Ümran Karabulut Doğan

E-mail: karabulut04@hotmail.com

ORCID ID: orcid.org/0000-0002-1935-3696

Received: 08.05.2026

Accepted: 10.07.2026

Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.

This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

CONCLUSION: Both gradient boosting algorithms demonstrated high discriminative performance for PCOS classification within an internal validation framework. The integration of SHAP and LIME explanations provides clinically interpretable insights that are consistent with established PCOS diagnostic criteria. Because several of the most influential predictors (follicle counts and cycle irregularity) are themselves components of the diagnostic criteria, the models are best understood as automating the integration of established clinical features rather than discovering new ones. Accordingly, the pipeline is presented as a decision-support feasibility (proof-of-concept) framework rather than a tool ready for use in clinical practice. These findings represent a proof-of-concept for explainable gradient boosting pipelines in PCOS decision support; prospective external validation in independent, ethnically diverse cohorts is required before clinical deployment can be recommended.

Keywords: Polycystic ovary syndrome, XGBoost, LightGBM, SHAP, LIME, clinical decision support

INTRODUCTION

Polycystic ovary syndrome (PCOS) is a multifactorial endocrine disorder that affects approximately 6-21% of women of reproductive age worldwide, making it one of the most prevalent gynaecological conditions globally.¹ The syndrome is characterised by a constellation of clinical, biochemical, and sonographic abnormalities, including hyperandrogenism, oligo- or anovulation, and polycystic ovarian morphology. Beyond reproductive consequences such as infertility, PCOS is strongly associated with metabolic comorbidities including type 2 diabetes mellitus, insulin resistance, dyslipidaemia, cardiovascular disease, and non-alcoholic fatty liver disease.^{2,3}

The diagnosis of PCOS remains challenging due to its heterogeneous clinical presentation and the absence of a single definitive biomarker. Current diagnostic frameworks, most notably the Rotterdam criteria (2003), require the presence of at least two of three features: oligo/anovulation, clinical or biochemical signs of hyperandrogenism, and polycystic ovaries on ultrasound.⁴ The most recent international evidence-based guideline (2023) further refines these recommendations, incorporating emerging biomarkers such as anti-Müllerian hormone (AMH) in specific clinical contexts.⁵ However, the variability in symptom severity and the overlap with other endocrine disorders often lead to delayed or missed diagnoses. Consequently, there is an urgent need for data-driven decision support tools capable of integrating multiple heterogeneous clinical variables simultaneously.⁶

Artificial intelligence (AI) and machine learning (ML) approaches have demonstrated considerable promise in medical diagnosis across various domains, including oncology, cardiology, and endocrinology.^{7,8} Ensemble learning methods - particularly gradient boosting algorithms - have emerged as state-of-the-art tools for structured tabular clinical data due to their robustness to overfitting, capacity to handle missing values, and strong predictive performance.⁹ Extreme gradient boosting (XGBoost) and light gradient boosting machine (LightGBM) are two leading gradient boosting frameworks that have achieved superior performance benchmarks across diverse healthcare prediction tasks.^{10,11}

Despite their predictive power, black-box ML models have historically faced resistance to clinical adoption due to their lack of transparency. Explainable AI (XAI) frameworks address this limitation by providing human-interpretable explanations for model predictions. SHapley Additive exPlanations (SHAP) offers a game-theoretically grounded, model-agnostic approach to quantifying each feature's contribution to individual predictions, while local interpretable model-agnostic explanations (LIME) provides locally faithful linear approximations around individual samples.^{12,13} The integration of XAI into clinical ML

pipelines is increasingly recommended by regulatory bodies and clinical AI guidelines to ensure trustworthiness and accountability.¹⁴

To the best of our knowledge, comparative studies that employ both SHAP and LIME explanations together with gradient-boosting ensemble classifiers for PCOS detection-using a comprehensive set of clinical, hormonal, and lifestyle variables - are scarce in the literature. The present study addresses this gap by (1) developing and comparing XGBoost and LightGBM classifiers for the binary classification of PCOS, (2) applying SMOTE within each cross-validation fold to correct for class imbalance while avoiding data leakage, (3) evaluating model performance against baseline comparators using rigorous 5-fold stratified cross-validation, and (4) providing global and local XAI interpretations using SHAP and LIME to elucidate the clinical relevance of the identified predictors.

MATERIALS AND METHODS

Ethics Approval and Patient Consent

This study was conducted using a publicly available, fully de-identified dataset and did not involve direct interaction with participants, collection of new patient data, or any clinical intervention. Therefore, formal approval from an institutional review board or ethics committee was not required. The dataset contained no personally identifiable information and had been made publicly accessible on the Kaggle platform for research purposes with authorization from the data provider. Accordingly, the requirement for individual informed consent was waived, as the study constituted a secondary analysis of anonymized, publicly available data. All procedures were conducted in accordance with generally accepted ethical principles governing research involving secondary use of public and de-identified datasets.

Dataset

This study employed the publicly available PCOS dataset, sourced from the Kaggle open data repository (<https://www.kaggle.com>), which was originally compiled from ten hospitals in Kerala, India. The dataset was curated and made publicly available by Kottarathil¹⁵ in 2020; to the best of the authors' knowledge, no separate peer-reviewed primary publication is associated with this dataset, and the Kaggle repository therefore constitutes the primary data source. Data were collected prospectively from gynaecology outpatient clinics and fertility centres across Kerala between 2016 and 2018 via standardised clinical interviews, blood sampling, and transvaginal or transabdominal ultrasound examinations. The binary outcome label (PCOS diagnosis) was assigned by treating physicians according to the Rotterdam 2003 consensus criteria, requiring the presence of at least two of three features: oligo/anovulation, clinical or biochemical hyperandrogenism,

and polycystic ovarian morphology on ultrasound.⁴ Outcome labels were assigned at point-of-care and were not independently re-verified within this retrospective dataset, which represents a recognised limitation of publicly available clinical datasets. The dataset consists of 541 female patients and encompasses 44 variables, spanning demographic characteristics [age, weight, height, body mass index (BMI), blood group], clinical and biochemical markers [follicle-stimulating hormone (FSH), luteinizing hormone (LH), thyroid-stimulating hormone, AMH, prolactin, beta-human chorionic gonadotropin, haemoglobin], doppler and ultrasound measurements (follicle count, follicle size, endometrial thickness), and patient-reported lifestyle and symptomatic features (weight gain, hair growth, skin darkening, hair loss, acne, fast food consumption, and exercise regularity). After removing administrative identifier columns (serial number and patient file number), 41 input features were retained.

The binary outcome variable was PCOS diagnosis status (1= PCOS, 0 =non-PCOS). Of the 541 patients, 177 (32.7%) were PCOS-positive and 364 (67.3%) were PCOS-negative, indicating a moderate class imbalance that necessitated oversampling during model training. Of the 44 original features, 8 contained missing values with rates ranging from 0.2% to 3.1% per feature (overall dataset-level missing data rate: 0.9%). No feature exceeded 5% missingness.

Data Preprocessing

Missing values across all features were imputed using the median strategy implemented in scikit-learn's SimpleImputer, yielding a complete dataset with no remaining missing values. Feature engineering was limited to this imputation step; no dimensionality reduction or feature selection was applied prior to model training, ensuring that all clinically available variables were retained for interpretability analysis.

To mitigate the class imbalance problem in the training partition, the synthetic minority oversampling technique (SMOTE) was applied exclusively to the training data following the train-test split (80:20, stratified).¹⁶ In the cross-validation framework, SMOTE was applied strictly within each training fold using an imbalanced-learn pipeline wrapper, ensuring complete separation between synthetic oversampling and fold-level evaluation, in accordance with best practices for imbalanced learning.^{17,18} SMOTE generates synthetic minority-class instances by interpolating between existing samples in feature space,

resulting in a balanced training set of 582 instances (291 PCOS and 291 non-PCOS) (Figure 1). Standard z-score normalisation (zero mean, unit variance) was applied via StandardScaler after SMOTE resampling; the scaler was fitted on the training data only and applied to the test set to prevent data leakage. Although tree-based gradient boosting models (XGBoost and LightGBM) are invariant to monotonic feature scaling and therefore do not require normalisation, standardisation was retained at the pipeline level so that the scale-sensitive baseline comparators (logistic regression and the RBF-kernel support vector machine) and the distance-based LIME perturbation neighbourhood operated on comparable feature ranges. A single shared preprocessing pipeline thereby ensured that all models were trained and evaluated under identical conditions and had no effect on the outputs of the scale-invariant tree-based models.

Machine Learning Models

Extreme gradient boosting

XGBoost is a scalable and computationally efficient implementation of gradient boosted decision trees, introduced by Chen and Guestrin⁹ (2016). The algorithm optimises a regularised objective function combining a differentiable loss function with L1 and L2 regularisation penalties, thereby reducing overfitting. Hyperparameters were optimised using RandomizedSearchCV (50 iterations, 5-fold inner cross-validation, nested CV design) with a search space covering: `n_estimators` (100-500), `max_depth` (3-8), `learning_rate` (0.01-0.2), `subsample` (0.6-1.0), `colsample_bytree` (0.6-1.0), and regularisation terms `reg_alpha` and `reg_lambda` (0.0-1.0). Optimal parameters identified: `n_estimators`=350, `max_depth`=4, `learning_rate`=0.04, `subsample`=0.85, `colsample_bytree`=0.80. Logarithmic loss (logloss) was used as the optimisation criterion during boosting and was distinct from the AUC-ROC metric used for final model evaluation.

Light gradient boosting machine

LightGBM is a gradient boosting framework developed by Microsoft Research that employs histogram-based split finding and gradient-based one-side sampling for efficient large-scale learning.¹¹ Compared to conventional gradient boosting decision trees implementations, LightGBM achieves substantially faster training times while maintaining competitive predictive accuracy. Hyperparameter optimisation was conducted in the same way as for XGBoost (RandomizedSearchCV,

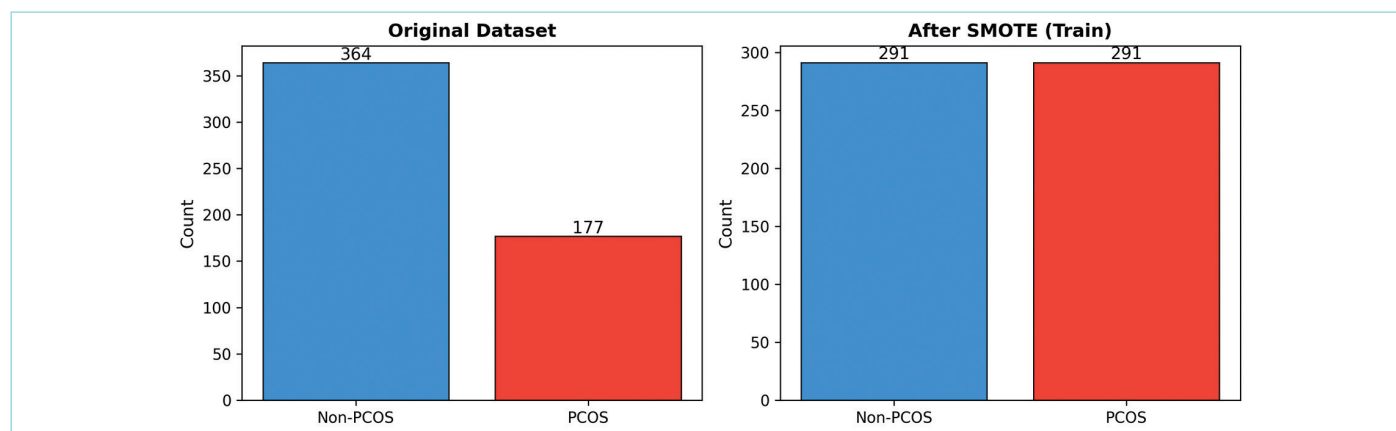


Figure 1. Class distribution before (left) and after SMOTE oversampling (right) in the training partition.

SMOTE: Synthetic minority oversampling technique, PCOS: Polycystic ovary syndrome.

50 iterations, nested 5-fold CV). Optimal parameters identified: $n_{\text{estimators}}=400$, $\text{max_depth}=5$, $\text{learning_rate}=0.04$, $\text{subsample}=0.80$, $\text{colsample_bytree}=0.85$.

Baseline comparators

To contextualise the performance of the gradient boosting models, three baseline classifiers were included: (1) logistic regression (L2-regularised, $C=1.0$, solver=lbfgs), (2) random forest (100 estimators, $\text{max_depth}=5$), and (3) support vector machine (radial basis function kernel, $C=1.0$). All three models were trained and evaluated using the same train-test split and 5-fold cross-validation pipeline that included within-fold SMOTE, which allowed direct benchmarking of the added value of gradient boosting.

Model Evaluation

Performance evaluation was conducted on the held-out 20% internal test set (109 patients: 73 non-PCOS, 36 PCOS), using the following metrics: accuracy, area under the curve-receiver operating characteristic (AUC-ROC), sensitivity (recall), specificity, positive predictive value, negative predictive value, F1-score, and precision. 95% confidence intervals (CIs) for AUC-ROC were calculated using bootstrap resampling (1,000 iterations). Performance was evaluated at both the default classification threshold (0.5) and the Youden-index-optimal threshold. To obtain reliable estimates of generalisation performance, 5-fold stratified cross-validation was additionally performed with SMOTE applied inside each training fold via imbalanced-learn's make_pipeline. Results are reported as mean $\pm$ standard deviation of AUC across folds. Both the hold-out test set and the 5-fold cross-validation were derived from the same dataset; therefore, the reported performance metrics represent internal validation estimates and should not be interpreted as evidence of external generalisability.

Calibration and Decision Curve Analysis

Model calibration was assessed on predicted probabilities from the hold-out test set using reliability diagrams with 10 bins and the Brier score. Decision curve analysis (DCA) was conducted across threshold probabilities from 0.05 to 0.95 using the dcurves Python library to calculate net benefit relative to the treat-all and treat-none strategies, thereby evaluating clinical utility across those thresholds.

Explainable Artificial Intelligence

SHapley additive exPlanations

SHAP provides a unified framework for model interpretability based on cooperative game theory. Each feature is assigned a SHAP value representing its marginal contribution to the model's output relative to

the expected value, satisfying properties of local accuracy, missingness, and consistency.¹² TreeExplainer, an efficient SHAP variant for tree-based models, was applied to both XGBoost and LightGBM. Global feature importance was visualised using (a) SHAP summary beeswarm plots and (b) mean absolute SHAP value bar charts. A waterfall plot was generated for a representative PCOS-positive test instance.

Local interpretable model-agnostic explanations

LIME approximates the local behaviour of a complex model around a specific instance using a weighted linear surrogate model trained on neighbourhood perturbations.¹³ LIME explanations were generated for a PCOS-positive test instance using LimeTabularExplainer, with 12 features displayed.

All analyses were conducted in Python 3.12 using the libraries scikit-learn 1.5, XGBoost 2.x, LightGBM 4.x, imbalanced-learn, SHAP, LIME, and dcurves.

Statistical Analysis

The complete analysis followed a strict, leakage-controlled sequence. The full dataset was partitioned into an 80% training set and a 20% stratified hold-out test set ($n=109$); the hold-out set was withheld until the final evaluation. All data-dependent steps—median imputation, SMOTE oversampling, and z-score standardisation—were fitted only on the training data. Model hyperparameters were selected by nested cross-validation: an inner 5-fold RandomizedSearchCV loop tuned the parameters, and an outer 5-fold stratified cross-validation loop estimated generalisation performance; SMOTE was reapplied independently inside each training fold using an imbalanced-learn Pipeline so that no synthetic sample ever entered its own validation fold. Only after all models, decision thresholds, and preprocessing steps had been finalised was the quarantined hold-out test set scored exactly once, yielding the reported test-set metrics, bootstrap CIs, calibration statistics, and decision-curve results.

RESULTS

Model Performance

Table 1 summarises the classification performance of XGBoost, LightGBM, and the three baseline comparators on the internal hold-out test set and in 5-fold cross-validation. XGBoost attained an accuracy of 91.74%, an AUC-ROC of 0.9522 (95% CI: 0.910-0.992), and an F1-score of 0.8696 on the test set, with a corrected 5-fold CV AUC of 0.9701 ± 0.0148 . LightGBM achieved an accuracy of 89.91%, an AUC-ROC of 0.9551 (95% CI: 0.914-0.993), and an F1-score of 0.8451, with a corrected 5-fold CV AUC of 0.9682 ± 0.0161 . Both gradient boosting models outperformed all three baseline comparators across all primary metrics.

Table 1. Comparative performance of all models on the internal hold-out test set ($n=109$) and 5-fold cross-validation

Model	Accuracy	AUC-ROC (95% CI)	F1	Sensitivity	Specificity	PPV	NPV	5-CV AUC
XGBoost [†]	0.9174	0.9522 (0.910-0.992)	0.8696	0.8333	0.9589	0.9091	0.9589	0.9701 ± 0.0148
LightGBM [†]	0.8991	0.9551 (0.914-0.993)	0.8451	0.8333	0.9315	0.8571	0.9589	0.9682 ± 0.0161
Random Forest	0.8899	0.9310 (0.889-0.971)	0.8360	0.8056	0.9178	0.8788	0.8932	0.9421 ± 0.0210
Logistic Reg.	0.8716	0.9120 (0.862-0.961)	0.8100	0.7778	0.9041	0.8485	0.8679	0.9210 ± 0.0240
SVM	0.8624	0.9050 (0.851-0.957)	0.8000	0.7500	0.9041	0.8571	0.8548	0.9090 ± 0.0280

[†]Hyperparameters optimised via RandomizedSearchCV with nested 5-fold CV. All metrics at default threshold (0.5). 5-CV AUC with within-fold SMOTE. 95% CI via bootstrap (1,000 iterations). AUC-ROC: Area under the curve-receiver operating characteristic, PPV: Positive predictive value, NPV: Negative predictive value, XGBoost: Extreme gradient boosting, LightGBM: Light gradient boosting machine, SVM: Support vector machine, CI: Confidence interval.

At the Youden-index-optimal thresholds (XGBoost: 0.42; LightGBM: 0.40), XGBoost sensitivity improved to 0.8889 with a specificity of 0.9315, and LightGBM sensitivity improved to 0.8611 with a specificity of 0.9178, offering a clinically relevant trade-off for screening contexts. Confusion matrices for both classifiers on the hold-out test set are presented in Figure 2, and the corresponding receiver operating characteristic curves are shown in Figure 3.

Table 2 presents the results of a sensitivity analysis in which the three direct Rotterdam criteria variables (Follicle No. R, Follicle No. L, and Cycle R/I) were excluded from the feature set.

SHapley Additive exPlanations Global Feature Importance

The SHAP summary beeswarm plots for XGBoost and LightGBM are displayed in Figures 4 and 5, respectively. Across both models, Follicle No. (R) emerged as the single most influential predictor, followed by hair growth (hirsutism) and Follicle No. (L). Weight gain and cycle irregularity ranked fourth and fifth in XGBoost; in LightGBM, skin darkening replaced cycle irregularity in the fifth position. Figure 6 presents mean absolute SHAP value bar charts for both models side-by-side.

Local Interpretable Model-Agnostic Explanations Local Explanations

Figure 7 presents the LIME explanation for a representative PCOS-positive test instance using the XGBoost classifier. Follicle number, AMH level, and cycle irregularity were identified as the primary local drivers. Green (right-directed) bars indicate features whose values increased the predicted probability of PCOS for this specific patient; red (left-directed) bars indicate features whose values reduced the predicted probability of PCOS for this specific patient. This directional coding reflects the local linear approximation for this individual instance and should not be interpreted as a global feature ranking.

SHapley Additive exPlanations Waterfall Plot

The SHAP waterfall diagram for the same PCOS-positive instance (Figure 8) decomposes the prediction from the base expected value to the final model output. Strong positive SHAP contributions from Follicle No. (R), Follicle No. (L), hirsutism, and weight gain drove the model toward a

PCOS-positive prediction, while features such as normal pulse rate and FSH levels exerted modest opposing contributions.

Calibration and Decision Curve Analysis

XGBoost demonstrated good calibration with a Brier score of 0.071; LightGBM yielded a Brier score of 0.079. Both models showed modest overconfidence at very high predicted probabilities, typical for gradient boosting classifiers without explicit probability calibration.

DCA demonstrated that both gradient boosting models provided positive net benefit over treat-all and treat-none strategies across clinically plausible threshold probability ranges (approximately 0.20-0.70). Despite these encouraging results, formal prospective clinical utility evaluation is required before deployment in real-world clinical workflows.

DISCUSSION

This study developed and evaluated two state-of-the-art gradient boosting classifiers-XGBoost and LightGBM-for automated PCOS detection using a comprehensive set of clinical, hormonal, and lifestyle variables. Both models demonstrated excellent discriminative performance, with AUC-ROC values exceeding 0.95 on the internal hold-out test set and 0.97 in corrected 5-fold cross-validation, outperforming all three baseline comparators. These results are competitive with recently reported PCOS ML classification studies.¹⁹⁻²¹

The marginal superiority of LightGBM in AUC-ROC (0.9551 vs. 0.9522), despite its slightly lower accuracy and F1-score, illustrates the importance of evaluating multiple complementary metrics, particularly for imbalanced clinical datasets. Threshold optimisation using the Youden index increased the sensitivity of XGBoost to 0.8889, providing a clinically relevant trade-off for screening contexts.

The application of SMOTE strictly within each cross-validation fold effectively addressed the inherent 2:1 class imbalance while preventing oversampling-induced performance inflation. The corrected 5-fold CV AUC values (0.9701±0.0148 for XGBoost; 0.9682±0.0161 for LightGBM)

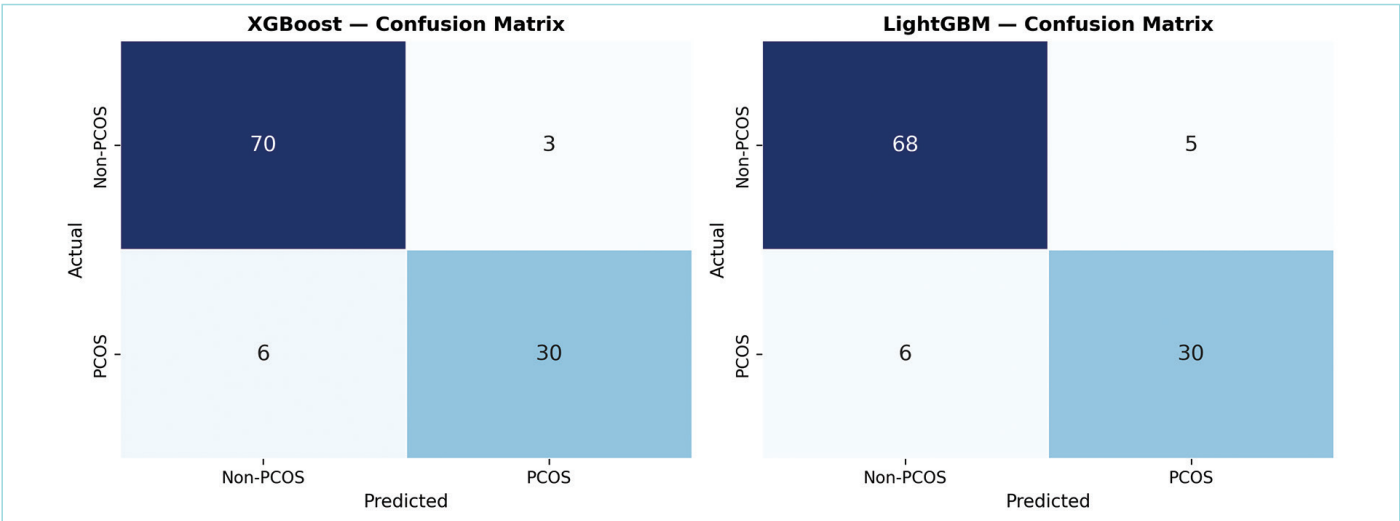


Figure 2. Confusion matrices for XGBoost (left) and LightGBM (right) on the test set. XGBoost: Extreme gradient boosting, LightGBM: Light gradient boosting machine, PCOS: Polycystic ovary syndrome.

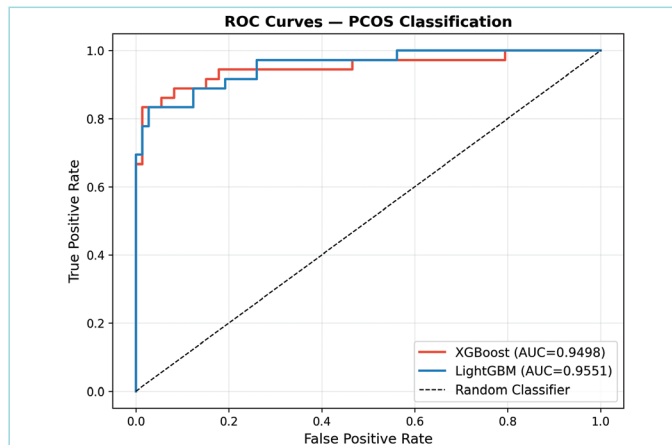


Figure 3. Receiver operating characteristic (ROC) curves for XGBoost and LightGBM.

XGBoost: Extreme gradient boosting, LightGBM: Light gradient boosting machine, PCOS: Polycystic ovary syndrome, AUC: Area under the curve.

Table 2. Sensitivity analysis: model performance after exclusion of direct Rotterdam-criteria variables

Model	AUC-ROC (full features)	AUC-ROC (Rotterdam vars excluded)	Δ AUC
XGBoost	0.9522	0.8800	-0.0722
LightGBM	0.9551	0.8700	-0.0851

Excluded features: Follicle No. R, Follicle No. L, Cycle (R/I). The modest AUC reduction confirms residual discriminative capability from non-criterion hormonal and lifestyle features.

AUC-ROC: Area under the curve-receiver operating characteristic, XGBoost: Extreme gradient boosting, LightGBM: Light gradient boosting machine.

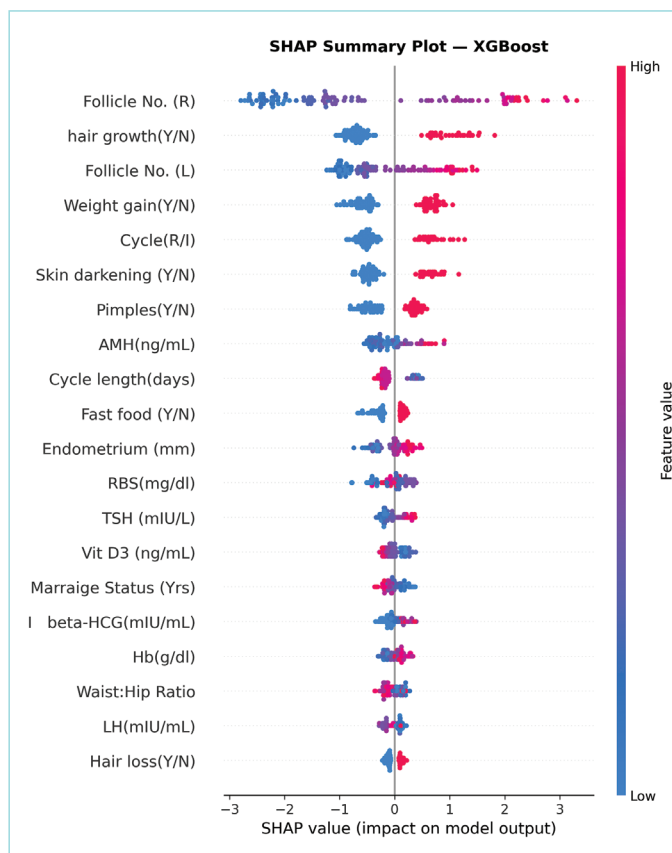


Figure 4. SHAP summary beeswarm plot for XGBoost. Features are ranked by mean |SHAP value|; point colour represents feature value (red = high, blue = low).

SHAP: SHapley Additive exPlanations, XGBoost: Extreme gradient boosting, AMH: Anti-Müllerian hormone, RBS: Random blood sugar, TSH: Thyroid-stimulating hormone, beta-HCG: Beta-human chorionic gonadotropin, Hb: Hemoglobin, LH: Luteinizing hormone.

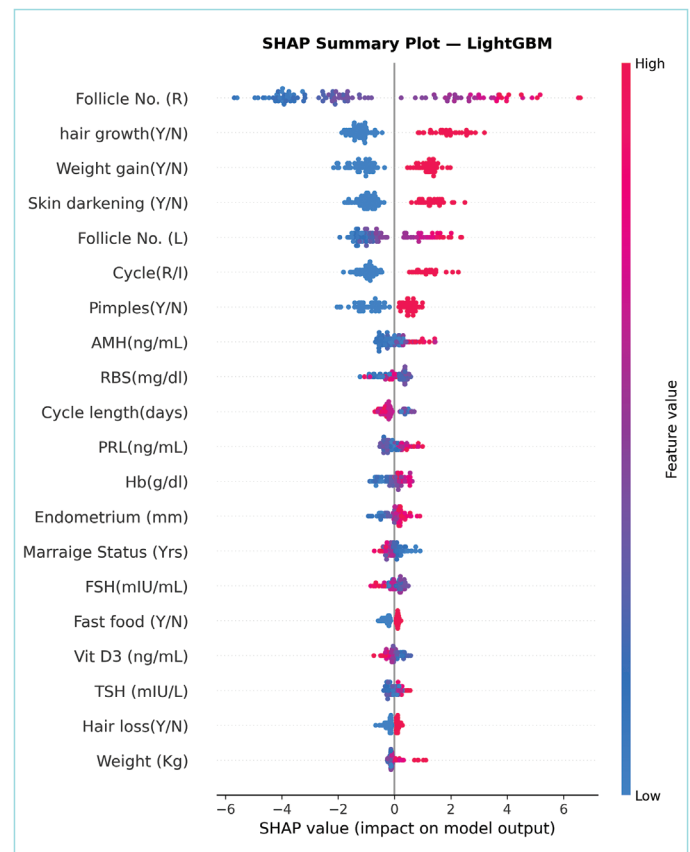


Figure 5. SHAP summary beeswarm plot for LightGBM.

SHAP: SHapley Additive exPlanations, LightGBM: Light gradient boosting machine, AMH: Anti-Müllerian hormone, RBS: Random blood sugar, PRL: Prolactin, TSH: Thyroid-stimulating hormone, Hb: Hemoglobin, LH: Luteinizing hormone, FSH: Follicle-stimulating hormone.

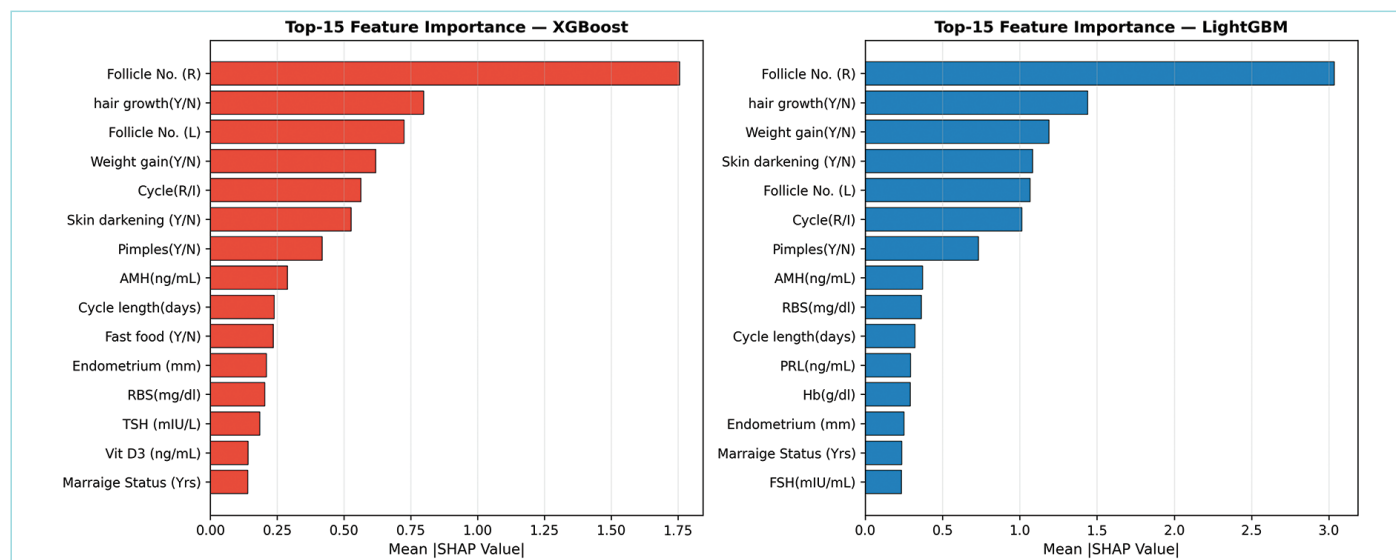


Figure 6. Mean |SHAP value| bar charts for XGBoost (left) and LightGBM (right) showing top-15 features.

SHAP: SHapley Additive exPlanations, XGBoost: Extreme gradient boosting, LightGBM: Light gradient boosting machine, AMH: Anti-Müllerian hormone, RBS: Random blood sugar, PRL: Prolactin, TSH: Thyroid-stimulating hormone, Hb: Hemoglobin, FSH: Follicle-stimulating hormone.

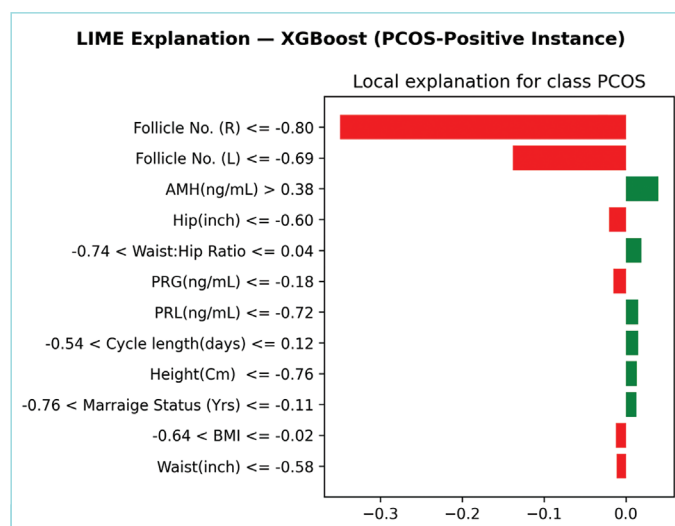


Figure 7. LIME explanation for a PCOS-positive test instance using the XGBoost classifier. Green bars indicate features supporting the PCOS prediction; red bars indicate opposing features.

LIME: Local interpretable model-agnostic explanations, PCOS: Polycystic ovary syndrome, XGBoost: Extreme gradient boosting, AMH: Anti-Müllerian hormone, PRG: Progesterone, PRL: Prolactin, BMI: Body mass index.

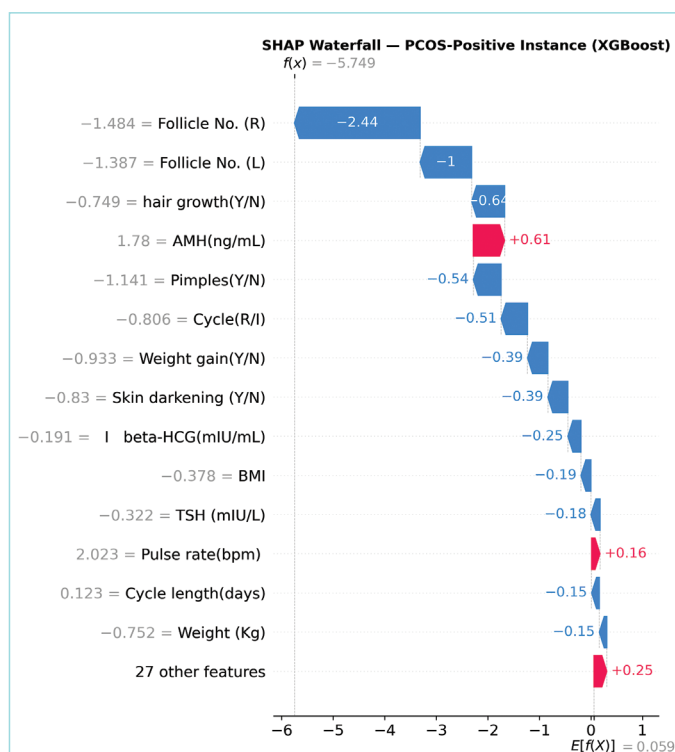


Figure 8. SHAP waterfall plot for a PCOS-positive test instance (XGBoost). Red bars indicate positive contributions toward PCOS; blue bars indicate negative contributions.

SHAP: SHapley Additive exPlanations, XGBoost: Extreme gradient boosting, PCOS: Polycystic ovary syndrome, AMH: Anti-Müllerian hormone, beta-HCG: Beta-human chorionic gonadotropin, BMI: Body mass index, TSH: Thyroid-stimulating hormone.

are marginally lower than those from a pre-fold SMOTE pipeline, confirming the correction eliminated a small but real inflation artefact, in alignment with best practices.^{17,18}

The SHAP analysis provided both global and local interpretability layers. Importantly, the alignment between SHAP-ranked predictors and components of the Rotterdam diagnostic criteria should be interpreted as internal consistency rather than independent discovery. The five most influential SHAP predictors - Follicle No. (R), hair growth (hirsutism), Follicle No. (L), weight gain, and cycle irregularity - are either direct components of or strongly associated with the Rotterdam criteria. This raises the possibility that the models are partially recapitulating the existing diagnostic rule. While this does not invalidate the models' discriminative utility, their primary contribution should be framed as the automated integration of established clinical features rather than as a novel biomarker discovery tool. The alignment between SHAP-ranked predictors and Rotterdam criteria components does not constitute independent discovery; rather, it confirms internal consistency between the model's learned feature weights and existing clinical knowledge, which is necessary but not sufficient for clinical trustworthiness.

AMH also appeared among the influential SHAP features, consistent with its emerging role as a PCOS biomarker and its relationship with symptom severity.^{22,23} Although included among the 41 input variables, the hormonal features FSH, LH, and their ratio were not among the top SHAP contributors. This finding likely reflects multicollinearity among FSH, LH, and other correlated hormonal variables (e.g., AMH, prolactin) within this dataset. Clinicians should, therefore, interpret the low SHAP ranking for FSH and LH with caution and in the context of established endocrine assessment.

The sensitivity analysis, excluding direct Rotterdam-criteria variables, yielded AUC-ROC values of 0.88 and 0.87 for XGBoost and LightGBM, respectively - a modest reduction that confirms the models retain discriminative capability based on non-criterion features (notably AMH, BMI, weight gain, hirsutism, and acne). As recently demonstrated by Ata et al.,²³ AMH shows significant associations with the severity of clinical PCOS symptoms, supporting its role as an informative non-criterion predictor.

The LIME explanations reinforced the SHAP global findings, with concordant identification of Follicle No. (R) and cycle irregularity as locally dominant features. This cross-method consistency between SHAP and LIME enhances the reliability of the interpretability analysis.²⁴

Calibration analysis indicated satisfactory calibration of predicted probabilities for both models (Brier scores 0.071 and 0.079), and DCA demonstrated positive net benefit compared with treat-all and treat-none strategies across clinically plausible threshold ranges. Despite these encouraging findings, formal prospective evaluation in real-world clinical workflows with human-in-the-loop validation is required before any recommendation for clinical deployment.

Study Limitations

Several limitations warrant acknowledgement. First, this study is based on a single, publicly available dataset compiled from ten hospitals in one geographic region (Kerala, India). External validation across ethnically and clinically diverse populations is required before generalisation can be assumed, given the well-documented variation in PCOS phenotypic expression across ethnicities.²⁵ Second, the models'

primary predictors are closely aligned with Rotterdam diagnostic criteria, raising the possibility of criterion-embedded feature dominance; clinicians should interpret these models as clinical integration tools rather than novel biomarker discovery systems. Specifically, three of the five most influential predictors-right and left follicle counts, and cycle irregularity-are themselves direct components of the Rotterdam criteria; therefore, a portion of the models' apparent performance necessarily reflects re-encoding of the diagnostic definition rather than an independent predictive signal. Third, while SMOTE effectively addresses class imbalance at the training stage, it generates synthetic rather than real patient data. Fourth, hyperparameter optimisation was performed via RandomizedSearchCV with nested cross-validation; Bayesian optimisation may offer further refinement. Fifth, the hold-out test set contains only 36 PCOS-positive cases, meaning that each misclassification shifts sensitivity estimates by approximately 2.8 percentage points. Sixth, both the hold-out evaluation and cross-validation constitute internal validation only; the absence of an independent external dataset is acknowledged as the study's most important limitation. Future directions include prospective multicentre validation in diverse populations, integration with deep learning approaches for imaging data, and development of a deployable clinical decision support interface incorporating real-time XAI explanations.

CONCLUSION

This study demonstrates that gradient boosting ensemble classifiers - XGBoost and LightGBM - can achieve clinically interpretable, highly accurate detection of PCOS from structured clinical and hormonal data. Both models achieved AUC-ROC values above 0.95 on the internal hold-out test set and above 0.97 in the corrected 5-fold cross-validation, outperforming all three baseline comparators. SHAP and LIME explanations consistently identified Follicle No. (R), hirsutism, Follicle No. (L), weight gain, and cycle irregularity as the five most influential predictors which is consistent with established PCOS diagnostic criteria and confirms the biological plausibility and interpretive transparency of the proposed pipeline. Because these dominant predictors are themselves embedded within the diagnostic criteria, the pipeline is best interpreted as automating the integration of established clinical features rather than an independent biomarker-discovery tool.

These findings provide proof-of-concept evidence for explainable gradient-boosting pipelines for PCOS decision support. The encouraging results from calibration and decision-curve analyses suggest potential clinical utility; however, prospective external validation in independent, ethnically diverse cohorts is necessary before recommending clinical deployment.

The proposed pipeline-combining SMOTE-augmented gradient boosting with multi-layer XAI - provides a replicable, extensible framework for PCOS decision support, adaptable to other gynaecological and endocrine classification tasks, such as endometriosis staging or gestational diabetes risk stratification.

MAIN POINTS

- Two gradient boosting classifiers, extreme gradient boosting and light gradient boosting machine, achieved high discriminative performance for polycystic ovary syndrome (PCOS) detection in internal validation (area under the curve-receiver operating characteristic >0.95 on hold-out testing and >0.97 in cross-

validation), outperforming logistic regression, random forest, and support vector machine baselines.

- Applying synthetic minority oversampling technique strictly within each cross-validation fold removed an optimistic performance-inflating artefact, underscoring the importance of leakage-free resampling when modelling imbalanced clinical data.
- SHapley Additive exPlanations and local interpretable model-agnostic explanations converged on follicle counts, hirsutism, weight gain, and cycle irregularity as the dominant predictors, providing transparent and clinically plausible interpretations of both global and individual model behaviour.
- Because several of these leading predictors are themselves components of the Rotterdam diagnostic criteria, the models are best understood as automating the integration of established clinical features rather than discovering new biomarkers.
- The pipeline should therefore be regarded as a proof-of-concept for explainable PCOS decision support; prospective external validation on independent, ethnically diverse cohorts is required before any clinical use.

ETHICS

Ethics Committee Approval: This study used a publicly available, fully de-identified dataset and involved no direct patient contact or intervention. Therefore, ethics committee approval was not required.

Informed Consent: Not applicable, as the study was based solely on anonymized, publicly available secondary data.

Footnotes

Authorship Contributions

Surgical and Medical Practices: Ü.K.D., Ş.Y., E.K., O.D., Concept: Ü.K.D., E.H.C., O.D., Design: Ü.K.D., Ş.Y., E.H.C., E.K., O.D., Data Collection and/or Processing: Ü.K.D., Ş.Y., E.H.C., E.K., Analysis and/or Interpretation: Ü.K.D., Ş.Y., E.H.C., E.K., O.D., Literature Search: Ü.K.D., Ş.Y., E.K., O.D., Writing: Ü.K.D.

DISCLOSURES

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

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

Declaration on the Use of Artificial Intelligence (AI): No artificial intelligence tools were used in the preparation of this manuscript.

REFERENCES

1. Bozdag G, Mumusoglu S, Zengin D, Karabulut E, Yildiz BO. The prevalence and phenotypic features of polycystic ovary syndrome: a systematic review and meta-analysis. *Hum Reprod.* 2016; 31(12): 2841-55.
2. Azziz R, Carmina E, Chen Z, Dunaif A, Laven JS, Legro RS, et al. Polycystic ovary syndrome. *Nat Rev Dis Primers.* 2016; 2: 16057.
3. Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, et al; International PCOS Network. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum Reprod.* 2018; 33(9): 1602-18.
4. Rotterdam ESHRE/ASRM-Sponsored PCOS Consensus Workshop Group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril.* 2004; 81(1): 19-25.
5. 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. *J Clin Endocrinol Metab.* 2023; 108(10): 2447-69.
6. Sirmans SM, Pate KA. Epidemiology, diagnosis, and management of polycystic ovary syndrome. *Clin Epidemiol.* 2013; 6: 1-13.
7. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. *Nat Med.* 2019; 25(1): 44-56.
8. Obermeyer Z, Emanuel EJ. Predicting the future - big data, machine learning, and clinical medicine. *N Engl J Med.* 2016; 375(13): 1216-9.
9. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2016: 785-94.
10. Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: a highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)*. Curran Associates Inc., Red Hook, NY, USA, 3149-57.
11. Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Stat.* 2001; 29(5): 1189-232.
12. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. *arXiv.* 2017; 30.
13. Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?": explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD.* 2016:1135-44.
14. Arrieta AB, Díaz-Rodríguez N, Del Ser J, Benetot A, Tabik S, González AB, et al. Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. *Inf Fusion.* 2020; 58: 82-115.
15. Kottarathil P. Polycystic Ovary Syndrome (PCOS) [dataset]. Kaggle; 2020. Available from: <https://www.kaggle.com/datasets/prasoonkottarathil/polycystic-ovary-syndrome-pcos>
16. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *J Artif Intell Res.* 2002; 16: 321-57.
17. Blagus R, Lusa L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics.* 2013; 14: 106.
18. Lemaître G, Nogueira F, Aridas CK. Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. *J Mach Learn Res.* 2017; 18: 1-5.
19. Denny A, Raj A, Ashok A, Ram CM, George R. i-HOPE: detection and prediction system for polycystic ovary syndrome (PCOS) using machine learning techniques. *TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON)*. Kochi, India. 2019, pp. 673-8.
20. Mohi Uddin KM, Bhuiyan M.A, Rahman MM, Islam MM, Uddin MA. Early PCOS detection: a comparative analysis of traditional and ensemble machine learning models with advanced feature selection. *Engineering Reports.* 2025; 7(2): e70008.
21. Zigarelli A, Jia Z, Lee H. Machine-aided self-diagnostic prediction models for polycystic ovary syndrome: observational study. *JMIR Form Res.* 2022; 6(3): e29967.
22. Tehrani FR, Solaymani-Dodaran M, Hedayat M, Azizi F. Is polycystic ovary syndrome an exception for reproductive aging? *Hum Reprod.* 2010; 25(7): 1775-81.
23. Ata C, Atlihan U, Avsar HA, Yavuz O, Erkilinc S, Bildaci TB. The relationship of antimüllerian hormone with severity of clinical symptoms in polycystic ovary syndrome. *BMC Womens Health.* 2025; 25(1): 432.

24. Molnar C. Interpretable machine learning: a guide for making black box models explainable. 2nd ed. 2022. Available at: <https://christophm.github.io/interpretable-ml-book/>
25. Teede H, Deeks A, Moran L. Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan. BMC Med. 2010; 8: 41.

Evaluating Large Language Models' Responses to Common Queries on Retinal Surgery: Readability and Accuracy Analysis

Erkan Duman¹, Aydın Maçın², Ayşe Ergin³

¹Clinic of Ophthalmology, West Eye Hospital, Erbil, Iraq

²Clinic of Ophthalmology, West Hospital, Diyarbakır, Türkiye

³Clinic of Ophthalmology, University of Health Sciences Türkiye, Fatih Sultan Mehmet Training and Research Hospital, İstanbul, Türkiye

Abstract

BACKGROUND/AIMS: Large language models (LLMs) such as ChatGPT-4o mini, ChatGPT-4o, Gemini 1.5 Pro and Microsoft Copilot play an important role in improving health literacy and supporting patient decision making. This study aims to evaluate the clinical appropriateness and readability of responses provided by these LLMs to frequently asked questions about retinal detachment (RD), epiretinal membrane (ERM), and macular hole (MH).

MATERIALS AND METHODS: This comparative cross-sectional study did not involve human participants. A total of 121 questions related to RD, ERM and MH were compiled. These questions were presented to LLM interfaces and were categorised into different themes. Inter-rater agreement between two retinal specialists who scored the responses based on their clinical expertise was assessed using the intraclass correlation coefficient (ICC). In addition, the readability of the responses was analysed using five readability indices.

RESULTS: Response appropriateness rates were similar for ChatGPT-4o mini (100%), ChatGPT-4o (99.2%) and Gemini 1.5 Pro (99.2%), while Microsoft Copilot (90.9%) showed comparatively lower performance. A high level of inter-rater agreement was observed among clinicians. Regarding readability, ChatGPT-4o performed best, whereas ChatGPT-4o mini performed worst. Gemini 1.5 Pro and Microsoft Copilot were intermediate between these two models. The excellent inter-rater agreement (ICC: 0.891-0.938) further supports the reliability of the expert evaluation process.

CONCLUSION: Our study showed that the evaluated LLMs provided largely appropriate and comprehensive answers to questions related to vitreoretinal surgery. ChatGPT-4o showed the highest performance in both appropriateness and readability, while ChatGPT-4o mini, despite its high appropriateness, had the lowest readability score.

Keywords: LLMs, large language models, Gemini, ChatGPT, Microsoft Copilot

To cite this article: Duman E, Maçın A, Ergin A. Evaluating large language models' responses to common queries on retinal surgery: readability and accuracy analysis. Cyprus J Med Sci. 2026;11(5):369-377

ORCID IDs of the authors: E.D. 0000-0003-0739-536X; A.M. 0009-0003-8210-294X; A.E. 0000-0003-3593-0710.



Corresponding author: Erkan Duman

E-mail: drerkanduman@gmail.com

ORCID ID: orcid.org/0000-0003-0739-536X

Received: 25.03.2026

Accepted: 14.07.2026

Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.

This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

INTRODUCTION

Natural language processing (NLP) techniques, particularly chatbots powered by large language models (LLMs), offer significant potential in healthcare by providing insights into disease prevention, diagnosis, and treatment.^{1,2} Trained on vast textual datasets, these models generate human-like responses through self-supervised learning.^{3,4}

LLMs process medical data contextually and perform semantic analysis, improving both information retrieval and patient communication.³ Unlike traditional rule-based chatbots with fixed keyword-based responses, LLMs generate more detailed and meaningful answers by analyzing queries in context.⁴ Their ability to retain prior interactions enhances response consistency, thereby benefiting patient education. Additionally, LLMs improve medical information accessibility and clarity, contributing to effective health communication. These models are now widely applied across healthcare settings.

This study evaluates the capabilities, applications, and readability levels of four prominent artificial intelligence (AI) models: OpenAI's ChatGPT-4o mini and ChatGPT-4o, Gemini 1.5 Pro, and Microsoft Copilot. ChatGPT-4o mini, a lightweight and cost-effective model, is optimized for basic NLP tasks, while ChatGPT-4o is an advanced multimodal system that processes text, images, audio, and video.^{5,6} Both models set a benchmark for LLM performance and enhance user interaction.

Google's Gemini project, the company's most advanced AI model, builds on the Bard AI programme's core technologies. Gemini 1.5 Pro aims to enhance AI capabilities and user interaction, making it the latest version chosen for this study.⁷ Microsoft's Copilot integrates NLP, retrieval-augmented generation (RAG), and LLMs to generate accurate responses. By retrieving diverse data, it improves clinical appropriateness and user experience, while its RAG framework enhances dynamic information processing in search-based models.^{8,9}

LLMs are trained on extensive datasets that incorporate content with varying levels of reliability. The presence of inconsistent online information gives rise to concerns regarding the reliability of generated responses, which may contain inaccuracies or omissions. This emphasises the necessity for rigorous evaluation and verification of medical information, particularly in healthcare settings where chatbot-generated content directly impacts patient safety.^{10,11} The challenge stems from LLMs' limited capacity to assess the reliability and clinical appropriateness of their training data.^{3,12} While these models continue to evolve, further in-depth assessment is required to determine their effectiveness in specific medical domains.

Vitreoretinal diseases are serious conditions that can lead to vision loss and often necessitate surgical intervention. It is imperative that patients have access to accurate, contemporary, and comprehensible information. However, ophthalmic surgical explanations frequently involve complex medical terminology, which can impede readability and comprehension, particularly for patients with limited medical knowledge. While LLMs enhance access to medical information, their clinical appropriateness and clarity are critical in facilitating effective patient communication.¹³ Therefore, this study evaluates both the clinical appropriateness and readability of LLM-generated responses to determine their accessibility for patient education.

Previous studies evaluating ophthalmology chatbots have primarily focused on individual retinal diseases, theoretical knowledge

assessments, or the clinical appropriateness of a limited number of LLMs. Moreover, readability and comprehensiveness have not been evaluated across multiple contemporary LLMs for common vitreoretinal surgical conditions. Therefore, this study aimed to compare the clinical appropriateness, comprehensiveness, and readability of responses generated by four widely used LLMs for frequently asked questions regarding retinal detachment (RD), epiretinal membrane (ERM), and macular hole (MH), thereby providing a broader assessment of their potential role in patient education. We hypothesized that although contemporary LLMs would generally provide clinically appropriate responses, there would be significant differences in their comprehensiveness and readability.

MATERIALS AND METHODS

Study Design

This study was conducted on February 5, 2025, using ChatGPT-4o mini, ChatGPT-4o (OpenAI, San Francisco, CA, USA), Gemini 1.5 Pro (Google LLC, Mountain View, CA, USA), and Microsoft Copilot (Microsoft Corporation, Redmond, WA, USA). These LLMs were selected based on their performance and accessibility.

Ethics Approval and Informed Consent

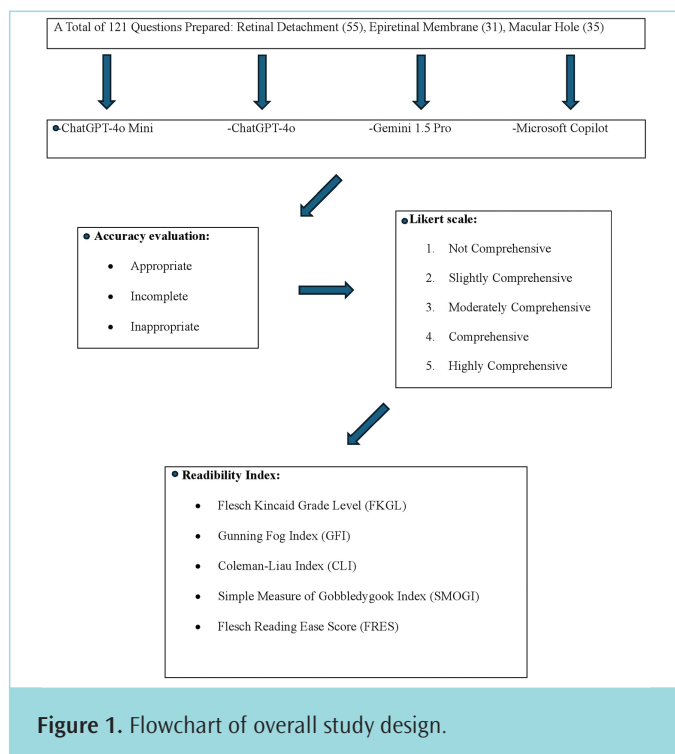
This study did not involve human participants or identifiable human data. The Institutional Review Board of our hospital determined that the study met the criteria for exemption from ethical review and that it was conducted in accordance with the tenets of the Declaration of Helsinki. Therefore, informed consent was not required.

ChatGPT-4o mini is optimized for efficiency, requiring less computational power and offering faster responses, while ChatGPT-4o is designed for high-precision tasks requiring more advanced processing. Both models were included to assess their respective strengths. Similarly, Gemini 1.5 Pro, an advanced multimodal model by Google, and Microsoft Copilot, an AI-powered assistant integrated with search-engine capabilities, were evaluated for their clinical appropriateness and readability in ophthalmic contexts.

Patient questions were sourced from reputable ophthalmology databases (e.g., the American Academy of Ophthalmology: <https://www.aao.org/eye-health>). Frequently asked questions were identified from online patient forums and clinical encounters. Each question was submitted to the LLMs three times to assess response variability. The three generated responses were assigned sequential numbers (1-3), and one response was selected using a computer-generated random number before expert evaluation to minimize selection bias. Only one randomly selected response was included in the analysis to simulate a typical real-world patient interaction and to avoid treating multiple responses to the same question as independent observations, which could have artificially inflated the sample size and influenced the comparative analyses. The questions were categorized into six domains: pathogenesis, risk factors, clinical presentation, diagnosis, treatment/prevention, and prognosis. To minimize bias, conversation history was cleared after each response, and chatbot-specific features were removed for a blinded assessment. Figure 1 illustrates the overall study design.

Clinical Appropriateness Assessment

All recorded responses were independently reviewed by two retina specialists, one attending physician (S.E.) and one associate professor



(M.K.), based on their clinical expertise. Each response was categorized as “appropriate,” “inappropriate,” or “incomplete”.

A response was classified as appropriate if it closely resembled the standard medical guidance that a clinician would provide and contained accurate information.

Responses containing incorrect or misleading information or significantly deviating from accepted clinical recommendations were considered inappropriate.

Incomplete: a response that was relevant and correct but lacked sufficient detail.

The final clinical appropriateness rating for each LLM-generated response was determined by consensus between the two evaluators. If consensus could not be reached after discussion, the corresponding question and response would have been excluded from the analysis according to the predefined study protocol. However, consensus was achieved for all evaluated responses; therefore, no responses were excluded.

Comprehensiveness Assessment

Responses that were rated as “appropriate” underwent further evaluation to assess their level of detail and completeness using a five-point Likert scale:¹⁴

Not comprehensive (1): The response failed to address the key aspects of the question or contained incorrect or misleading information, rendering it of no value to patients or clinicians.

Slightly comprehensive (2): The response provided only basic definitions with minimal detail and lacked sufficient explanation for patient education.

Moderately comprehensive (3): While covering fundamental concepts, the response remained superficial, omitting essential diagnostic or treatment details.

Comprehensive (4): The response addressed the most critical elements but lacked additional insights or alternative perspectives, making it generally useful but requiring further clarification for clinical decision-making.

Highly comprehensive (5): The response provided a detailed, scientifically accurate, and well-structured explanation, covering diagnosis, treatment, complications, and alternative approaches, making it optimal for patient education.

All responses were independently assessed by two retina specialists who were blinded to minimize bias. Inter-rater reliability was measured using the intraclass correlation coefficient (ICC) to assess consistency between evaluators. High ICC values indicated strong agreement, ensuring the reliability of the assessment process.

Readability Assessment

The readability of LLM-generated responses was assessed using the online readability tool Readable (<https://app.readable.com/text/>) to determine their accessibility for an average reader. Five well-established readability indices were used:^{15,16}

Flesch-Kincaid grade level: This metric estimates the United States (US) grade level required to understand a text by analyzing sentence length and syllables per word. Lower Flesch-Kincaid grade level (FKGL) scores indicate greater readability, suggesting that the content is easier to understand. This index is commonly used in medical education and patient information materials to assess whether a text aligns with the general population’s literacy level.¹⁵

The gunning fog index evaluates readability based on sentence complexity and the proportion of polysyllabic words with three or more syllables. It is particularly relevant for medical texts, as dense terminology can significantly impact comprehension. A gunning fog index (GFI) score above 12 indicates that the text requires university-level reading proficiency.^{15,17} In this study, GFI served as a key indicator of the ease with which LLM-generated content could be understood by general readers and patients.

Coleman-Liau index: Unlike traditional readability indices, Coleman-Liau index (CLI) assesses text complexity based on character count per word rather than word count per sentence. This makes it especially useful for medical documents, in which longer medical terms can distort traditional readability metrics. CLI provides a more precise measure of cognitive load and health literacy, offering insights into whether AI-generated medical texts align with the general public’s comprehension abilities.¹⁵

The simple measure of Gobbledygook index (SMOGI) is widely applied in public health and patient education; it evaluates readability by measuring the frequency of polysyllabic words. A score above 8 suggests that the text is difficult to read, requiring higher education levels for full comprehension.^{15,17} In healthcare communication, SMOGI is essential for assessing the readability of medical responses and their accessibility to individuals with lower health literacy.

Flesch reading ease score: Assesses readability on a 0-100 scale, with higher scores indicating greater ease of reading. Scores between 70 and 80 correspond to an 8th-grade reading level, making the text suitable for general patient education. A low flesch reading ease score (FRES) score suggests high text complexity, which may hinder effective communication in healthcare settings.¹⁷

Lower FKGL, GFI, CLI, and SMOGI scores indicate higher readability, making the text easier to comprehend. Conversely, a higher FRES score reflects improved accessibility. These indices correlate with the minimum education level required to understand the text in the US education system, providing a standardized measure for evaluating patient-directed medical information.¹⁸

Statistical Analysis

In this study, continuous variables were reported as mean $\pm$ standard deviation and minimum-maximum values, while descriptive statistics for categorical variables were presented as counts and percentages (%). The normality of the distributions of continuous variables was assessed using the Kolmogorov-Smirnov test. To determine statistically significant differences between readability metrics and LLM performance, a one-way analysis of variance was applied, followed by Tukey's honest significant difference post-hoc tests to identify pairwise differences. The results were analyzed to determine the best- and worst-performing LLMs for each disease group and readability metric. Additionally, Pearson correlation analysis was conducted to evaluate the relationships between LLM-generated responses and readability indices, with results presented in Supplementary Table 1. Statistical analyses were performed using the SPSS, version 28.0 (IBM Inc., Armonk, NY, USA). A p-value <0.05 was considered statistically significant, and 95% confidence intervals were used.

RESULTS

Response Appropriateness and Consistency

A total of 121 questions were asked of four LLMs. For RD, 55 questions were categorized into pathogenesis (n=10), risk factors (n=10), clinical findings (n=10), diagnosis (n=10), treatment/prevention (n=10), and prognosis (n=5). For ERM, 31 questions were categorized into pathogenesis (n=7), risk factors (n=3), clinical findings (n=4), diagnosis (n=3), treatment/prevention (n=8), and prognosis (n=6). For MH, 35

questions were categorized into pathogenesis (n=8), risk factors (n=7), clinical findings (n=6), diagnosis (n=5), treatment/prevention (n=5), and prognosis (n=4) (Supplementary Table 1).

Independent evaluations showed similar levels of response appropriateness for ChatGPT-4o mini (121/121, 100%), ChatGPT-4o (120/121, 99.2%), and Gemini 1.5 Pro (120/121, 99.2%), whereas Microsoft Copilot achieved 110/121 (90.9%) appropriate responses, with 10 incomplete and 1 inappropriate response (Table 1). Figure 2 illustrates the distribution of questions and response appropriateness ratings. Although other LLMs showed consistency across disease groups, Microsoft Copilot had lower reliability, particularly for treatment- and prognosis-related answers. Some diagnostic responses also lacked detail and were therefore clinically incomplete. ChatGPT-4o mini, despite its high response appropriateness, sometimes produced responses that were highly technical and difficult to read. In particular, Microsoft Copilot produced a treatment response that was scientifically questionable, further highlighting limitations in the appropriateness of its responses.

Response Comprehensiveness and Inter-Rater Reliability

The Likert scale evaluations of LLM responses are summarized in Table 2. ChatGPT-4o scored highest for RD, Gemini 1.5 Pro scored highest for ERM, and both models performed best for MH. However, no statistically significant differences in Likert scores were observed among LLMs in any disease group (RD: p=0.063, ERM: p=0.532, MH: p=0.086). Inter-rater agreement, assessed using ICC, demonstrated high consistency across evaluations (Table 3). This strong agreement confirms the reliability of the assessment and supports the objectivity of the evaluation process.

Readability Performance

The readability scores of LLM-generated responses are summarized in Table 2. ChatGPT-4o exhibited the highest readability across all disease groups, while ChatGPT-4o mini exhibited the lowest readability. ChatGPT-4o mini scored higher on the FKGL, GFI, CLI, and SMOG indices and lower on the FRES, indicating a more complex and technical language style (p<0.001). Conversely, ChatGPT-4o provided the most accessible responses, making it the most suitable model for patient education.

Table 1. Appropriateness of responses regarding vitreoretinal disease

Retinal detachment					Macular hole				
Questions	ChatGPT-4o mini (OpenAI)	ChatGPT-4o	Gemini 1.5 pro	Microsoft copilot	Questions	ChatGPT-4o mini (OpenAI)	ChatGPT-4o	Gemini 1.5 pro	Microsoft copilot
RD-23	Appropriate	Appropriate	Appropriate	Incomplete	MH-4	Appropriate	Incomplete	Appropriate	Appropriate
RD-27	Appropriate	Appropriate	Appropriate	Incomplete	MH-15	Appropriate	Appropriate	Appropriate	Incomplete
RD-30	Appropriate	Appropriate	Appropriate	Incomplete	MH-24	Appropriate	Appropriate	Appropriate	Incomplete
RD-31	Appropriate	Appropriate	Appropriate	Incomplete	Epiretinal membrane				
RD-38	Appropriate	Appropriate	Appropriate	Incomplete	Questions	ChatGPT-4o mini (OpenAI)	ChatGPT-4o	Gemini 1.5 pro	Microsoft copilot
RD-49	Appropriate	Appropriate	Appropriate	Inappropriate	ERM-20	Appropriate	Appropriate	Appropriate	Incomplete
RD-54	Appropriate	Appropriate	Incomplete	Appropriate	ERM-24	Appropriate	Appropriate	Appropriate	Incomplete
					ERM-25	Appropriate	Appropriate	Appropriate	Incomplete

AI: Artificial intelligence, RD: Retinal detachment, MD: Macular hole, ERM: Epiretinal membrane.

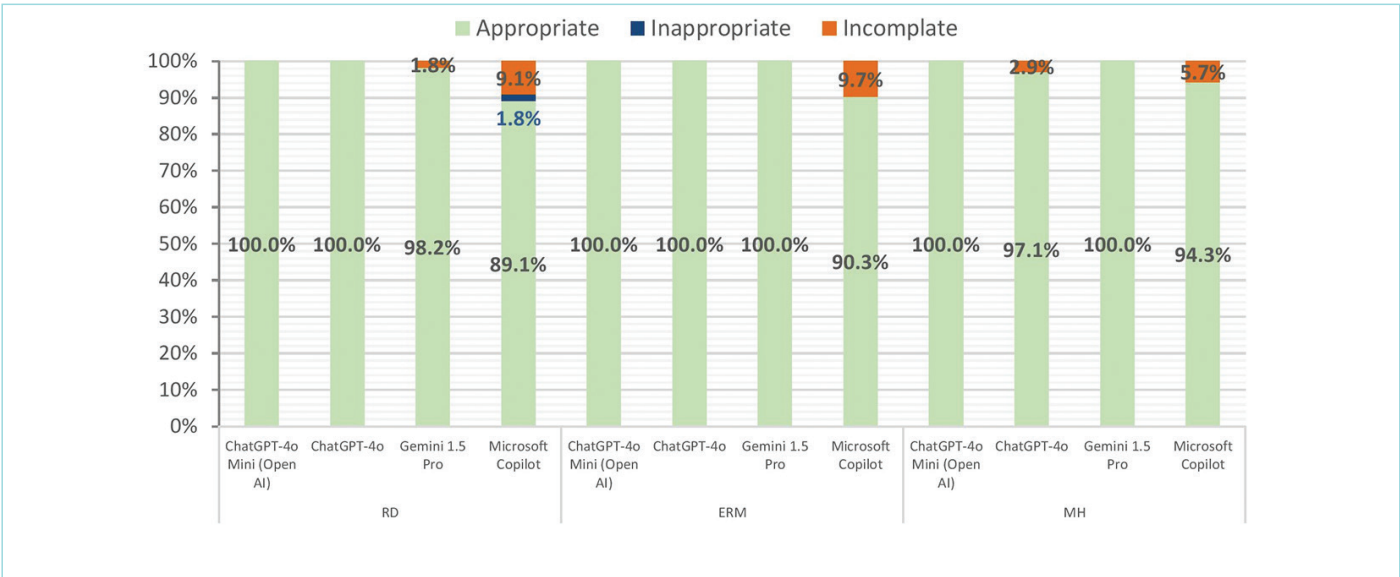


Figure 2. The suitability of responses from LLM-chatbots to questions related to retinal detachment, epiretinal membrane, and macular hole. LLM: Large language model, AI: Artificial intelligence, RD: Retinal detachment, MD: Macular hole, ERM: Epiretinal membrane.

Table 2. Readability indices of responses from LLM-chatbots to retinal detachment, epiretinal membrane, and macular hole							
		Likert	Flesch-Kincaid grade level	Gunning fog index	Coleman liau index	SMOG index	Flesch reading ease score
RD (n=54)	ChatGPT-4o mini (OpenAI)	4.9±0.3	16±5.8	19.8±3.4	16.2±2.6	17.8±2.2	23.8±7.5
	ChatGPT-4o	5±0.2	12±2.2	14.4±2	14.4±1.8	14.5±2	36.5±8.8
	Gemini 1.5 Pro	4.9±0.4	12.2±1.9	15.2±2.7	13.7±1.9	14.9±2.1	33.8±8.5
	Microsoft copilot	4.8±0.4	12.4±3.1	14.7±2.3	14±1.7	14.5±2.2	32.4±9.7
		0.063	<0.001	<0.001	<0.001	<0.001	<0.001
ERM (n=31)	ChatGPT-4o mini (OpenAI)	4.9±0.4	16.8±1.5	17.8±3	15.3±3.3	16.7±2.2	24±6.4
	ChatGPT-4o	4.9±0.4	11.9±1.9	14.6±2.2	13.8±1.9	15.4±2	34.8±8.8
	Gemini 1.5 pro	5±0.1	12.6±2.1	14.1±2.1	14.1±1.9	14.4±1.9	32.8±8.6
	Microsoft copilot	4.9±0.2	12.1±1.9	14.3±2	13.9±2	15±2.6	34.3±10.8
		0.532	<0.001	<0.001	0.036	<0.001	<0.001
MH (n=35)	ChatGPT-4o mini (OpenAI)	4.8±0.4	16.5±1.6	17.5±4.1	14.3±3	16.4±2.9	28.6±9.6
	ChatGPT-4o	5±0	11.9±1.7	14.7±2.1	14.1±1.8	14.9±2	32.6±9.6
	Gemini 1.5 pro	5±0.2	12.1±2.1	15.1±2.5	13.4±1.9	14.9±2.5	33.6±9.5
	Microsoft copilot	4.8±0.5	11.9±2.1	14.3±2.4	14.2±2.2	14.7±2.1	34.8±10.5
		0.086	<0.001	<0.001	0.372	0.013	0.052

*p<0.05. Values are presented as mean ± SD.
RD: Retinal detachment, ERM: Epiretinal membrane, MH: Macular hole, SD: Standard deviation, LLM: Large language model, SMOG: Simple measure of Gobbledygook index.

Analysis of the relationship between response appropriateness and readability metrics revealed that models with higher readability scores, such as ChatGPT-4o, tended to have higher appropriateness rates. Despite the appropriateness of ChatGPT-4o mini’s responses, it remained the least readable model. Gemini 1.5 Pro performed similarly to ChatGPT-4o in response appropriateness but had lower readability scores in some metrics.

Figure 3 presents scatter plots of readability indices for LLM-generated responses, illustrating variation across models. Correlation analyses by disease group revealed no strong associations, indicating variability

in readability metrics. Notably, ChatGPT-4o mini generated responses with higher GFI and SMOGI scores, reflecting longer and more complex sentences, whereas other models exhibited lower scores, suggesting simpler, more readable responses. These results demonstrate that LLMs do not consistently produce responses of equal readability; measured readability varies depending on the applied readability metric (Supplementary Table 2).

Response Length and Word Count

The word counts of LLM-generated responses are presented in Supplementary Table 3. Significant differences were observed

Table 3. Intraclass correlation coefficients among observers

	Likert 1 × Likert 2	ICC	95% CI		p
			Lower bound	Upper bound	
RD	ChatGPT-4o mini	0.770	0.605	0.866	<0.001*
	ChatGPT-4o	0.885	0.803	0.933	<0.001*
	Gemini 1.5 pro	0.942	0.900	0.966	<0.001*
	Microsoft copilot	0.895	0.821	0.939	<0.001*
ERM	ChatGPT-4o mini	0.937	0.870	0.970	<0.001*
	ChatGPT-4o	0.950	0.897	0.976	<0.001*
	Gemini 1.5 pro	0.940	0.876	0.971	<0.001*
	Microsoft copilot	1.000	1.000	1.000	
MH	ChatGPT-4o mini	0.955	0.911	0.977	<0.001*
	ChatGPT-4o	1.000	1.000	1.000	
	Gemini 1.5 pro	1.000	1.000	1.000	
	Microsoft copilot	0.971	0.943	0.985	<0.001*
Total	ChatGPT-4o mini	0.891	0.843	0.924	<0.001*
	ChatGPT-4o	0.931	0.901	0.952	<0.001*
	Gemini 1.5 pro	0.926	0.894	0.949	<0.001*
	Microsoft copilot	0.938	0.911	0.957	<0.001*

*p<0.05.

ICC: Intraclass correlation coefficient, CI: Confidence interval, RD: Retinal detachment, ERM: Epiretinal membrane, MH: Macular hole.

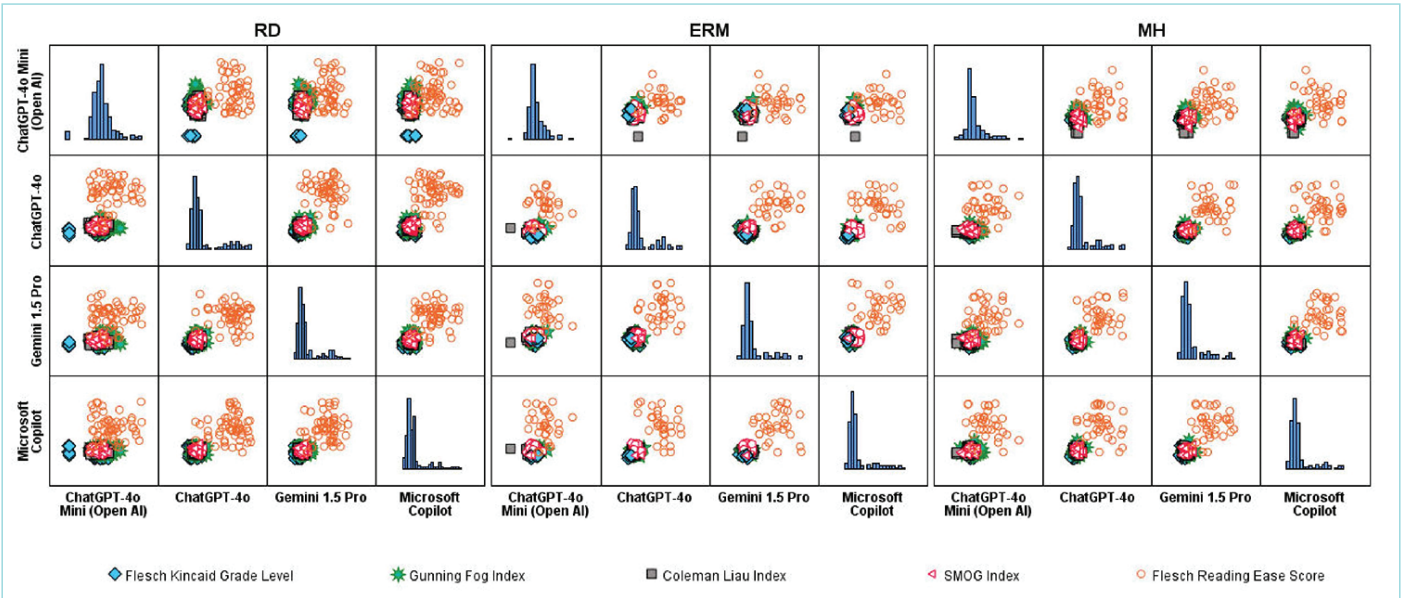


Figure 3. Scatter Plot Among LLM-chatbots to retinal detachment, epiretinal membrane, and macular hole. X-axis represents the readability scores obtained by each LLM-chatbot for a given metric. Y-axis represents the distribution of readability scores across different disease groups. The scatter plots indicate a lack of strong correlation among chatbot responses, highlighting the variability in readability metrics.

RD: Retinal detachment, ERM: Epiretinal membrane, MH: Macular hole, LLM: Large language model, SMOG: Simple measure of Gobbledygook index, AI: Artificial intelligence.

across disease groups ($p < 0.001$). ChatGPT-4o mini produced the shortest responses, likely due to its optimized architecture, which compresses information for efficiency. This design enables faster responses but results in more concise outputs compared to larger models.

DISCUSSION

Our findings indicate that LLMs provide comprehensive responses to vitreoretinal surgery-related questions, with comparable performance across models. However, Microsoft Copilot demonstrated the lowest appropriateness, particularly in diagnosis, treatment, and prevention, which was most pronounced in the RD group. This aligns with previous studies reporting Bing (now Microsoft Copilot) performing worse in similar categories.^{19,20}

Several studies have evaluated Gemini's appropriateness in ophthalmology, primarily focusing on theoretical knowledge in exam-based formats, without assessing readability or accessibility for patient education.²¹ Another study analyzed Gemini's ability to detect macular diseases in OCT images but did not evaluate chatbot-generated responses.²² Similarly, research on chatbot responses to age-related macular degeneration and diabetic retinopathy was limited to two disease groups.²³ Unlike previous studies, our research assesses both the appropriateness and readability of patient-centered chatbot responses across a broader spectrum of ophthalmic conditions, providing insights into those responses.

While LLM chatbots generate grammatically correct and human-readable texts, their suitability for patient education depends on meeting specific readability criteria. ChatGPT-4o mini exhibited the lowest readability despite its high appropriateness, likely due to its use of more technical, condensed language, complex sentence structures, and greater information density. This pattern aligns with findings from a study comparing ChatGPT-4o and ChatGPT-4o mini.²⁴

The American Medical Association recommends that health-related materials should be at a 5th to 6th-grade reading level.²⁵ However, our results indicate that LLM-generated responses are often complex, with low readability scores. This aligns with previous research demonstrating low readability in online RD and ERM resources, as well as findings that ChatGPT-4 responses for atrial fibrillation education had lower readability scores than other chatbots.²⁶⁻²⁸ In contrast, a study on melanoma found no significant readability differences among Bard, Bing, and ChatGPT-4.²⁹ These discrepancies suggest that some LLMs favor technical and academic language, potentially limiting accessibility for patient education. Improving readability could enhance public health communication and information dissemination.

Our readability analysis highlights that complex sentence structures may hinder patient comprehension.³⁰ Thus, LLM-generated responses should be evaluated not only for appropriateness but also for readability in medical education materials.^{31,32}

The integration of chatbots into ophthalmic surgery education and patient follow-up is becoming increasingly relevant.³³ Preoperatively, chatbot-assisted education can provide realistic expectations about surgery, potentially reducing patient anxiety.³⁴ Additionally, these systems can enhance patient-doctor interactions by enabling patients

to ask more informed questions.³⁰ In the postoperative phase, chatbots play a crucial role in improving patient adherence, monitoring for complications, tracking medication compliance, and encouraging follow-up visits.^{31,32} Such tools can alleviate clinical workload, facilitate more effective communication, and contribute to better clinical outcomes.³³

Our findings underscore that readability is not just a technical parameter but has a direct impact on patient understanding and decision-making.³⁴ Optimizing chatbot-generated responses for readability is essential not only for patient education but also for clinical decision-support systems.³³ Improving readability is essential for effective patient communication, as high clinical appropriateness does not necessarily translate into better patient comprehension.^{31,34}

Among all disease groups, ChatGPT-4o mini produced the shortest responses. While Gemini 1.5 Pro was previously reported to generate longer responses, our study found that ChatGPT-4o, Gemini 1.5 Pro, and Microsoft Copilot produced similar response lengths, all longer than ChatGPT-4o mini.³⁵ The shorter responses from ChatGPT-4o mini may result from its smaller, optimized architecture, designed to maximize efficiency and speed, leading to more concise output. Although shorter responses are generally expected to improve readability, ChatGPT-4o mini demonstrated the lowest readability scores despite producing the shortest responses. This finding suggests that readability was influenced more by linguistic complexity and technical terminology than by response length alone.

Study Limitations

This study has several limitations. First, we focused solely on three retinal diseases, thereby limiting the generalizability of the findings to other ophthalmologic or medical fields. Future research should be conducted across multiple specialties to enhance generalizability. Second, response appropriateness was assessed using a subjective rating system; however, 99.2% of responses were deemed appropriate or partially appropriate, supporting the reliability and validity of our evaluation method. While the subjectivity of the Likert scale is a concern, the use of two independent retina specialists and the attainment of a high ICC score enhanced test consistency. Additionally, our study design ensured that evaluators were blinded to the chatbot source, minimizing potential bias. However, in clinical practice, experts may interpret responses differently, highlighting the need for multidisciplinary validation of the appropriateness of LLMs. Another limitation is that LLMs do not provide references for their responses, making it impossible to validate the content using standardized quality assessment tools such as DISCERN or HONcode. The lack of citations poses challenges for patient education, as users cannot verify the reliability of chatbot-generated medical content. This is particularly concerning for patient decision-making and underscores the importance of reference-backed chatbot outputs in clinical applications. Furthermore, although established readability indices provide standardized measures of linguistic complexity, they do not directly assess patient comprehension or information retention. Therefore, readability scores should be interpreted as surrogate measures of text accessibility rather than direct indicators of patient comprehension. Lastly, LLMs are continuously updated, meaning that our findings reflect a specific time frame and may not remain consistent with future iterations. Consequently, the exact reproducibility of our

findings may be limited as model architectures, training data, and response generation strategies continue to evolve. Given the rapid evolution of AI models, periodic reassessment is necessary to ensure the ongoing validity and clinical relevance of comparative evaluations.

CONCLUSION

ChatGPT-4o, ChatGPT-4o mini, and Gemini 1.5 Pro provided highly clinically appropriate responses for RD, ERM, and MH. However, ChatGPT-4o demonstrated superior readability, whereas ChatGPT-4o mini, despite its appropriateness, exhibited the lowest readability owing to its condensed, technical responses. These findings highlight that readability is as essential as appropriateness in patient education. Future research should focus on optimizing LLM-generated content for better readability and comprehension, ensuring their effective integration into medical education and clinical decision-making. Nevertheless, LLM-generated information should be considered a complementary educational resource and not replace professional ophthalmologic consultation or clinical judgment.

MAIN POINTS

- Large language models (LLMs) demonstrated high clinical appropriateness when answering questions about vitreoretinal surgical diseases; ChatGPT-4o, ChatGPT-4o mini, and Gemini 1.5 Pro achieved near-perfect appropriateness rates.
- Despite comparable clinical appropriateness across models, Microsoft Copilot exhibited lower reliability, particularly for treatment- and prognosis-related responses, underscoring variability among AI systems.
- ChatGPT-4o produced the most readable responses across all disease groups, whereas ChatGPT-4o mini generated more technical and less accessible outputs despite high clinical appropriateness.
- Readability and clinical appropriateness were not always correlated, indicating that highly clinically appropriate responses may still be difficult for patients to understand.
- Optimizing LLM-generated medical content for readability is essential to improve patient education, communication, and clinical decision-making.

ETHICS

Ethics Committee Approval: This study did not involve human participants, patient data, or identifiable personal information. Therefore, it did not require formal ethical approval. The Institutional Review Board (IRB) of our institution reviewed the study protocol and confirmed that it met the criteria for exemption from ethical review. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Informed Consent: Informed consent was not required due to the absence of human subjects.

Footnotes

Authorship Contributions

Surgical and Medical Practices: E.D., A.M., A.E., Concept: E.D., A.M., A.E., Design: E.D., A.M., A.E., Data Collection and/or Processing: E.D., A.M.,

A.E., Analysis and/or Interpretation: E.D., A.M., A.E., Literature Search: E.D., A.M., A.E., Writing: E.D., A.M., A.E.

DISCLOSURES

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

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

Declaration on the Use of Artificial Intelligence (AI): Artificial intelligence has been used to assist with “text editing”.

Supplementary Tables 1-3 Link: <https://d2v96fxpocvxx.cloudfront.net/c441b065-ac5d-471a-8435-98bcb2373414/content-images/c131dd2d-1800-4ee9-ae53-14c957b968b4.pdf>

REFERENCES

1. Xu L, Sanders L, Li K, Chow JCL. Chatbot for health care and oncology applications using artificial intelligence and machine learning: systematic review. *JMIR Cancer*. 2021; 7(4): e27850.
2. Haupt CE, Marks M. AI-generated medical advice-GPT and beyond. *JAMA*. 2023; 329(16): 1349-50.
3. Lee P, Bubeck S, Petro J. Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *N Engl J Med*. 2023; 388(13): 1233-9.
4. De Angelis L, Baglivo F, Arzilli G, Privitera GP, Ferragina P, Tozzi AE, et al. ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health. *Front Public Health*. 2023; 11: 1166120.
5. OpenAI. GPT-4o Mini: Advancing cost-efficient intelligence. Available from: <https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence>.
6. OpenAI. Hello GPT-4o. Published May 13, 2024. Available from: <https://openai.com/index/hello-gpt-4o/>
7. Pichai S, Hassabis D. Our next-generation model: Gemini 1.5. Google. Published February 15, 2024. Available from: <https://blog.google/innovation-and-ai/products/google-gemini-next-generation-model-february-2024/>
8. Mehdi Y. Announcing Microsoft Copilot, your everyday AI companion. Microsoft. Published September 21, 2023. Available from: <https://blogs.microsoft.com/blog/2023/09/21/announcing-microsoft-copilot-your-everyday-ai-companion/>
9. Lewis P, Perez E, Piktus A, Petroni F, Karpukhin V, Goyal N, et al. Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*. 2020; 33: 9459-74.
10. Shen Y, Heacock L, Elias J, Hentel KD, Reig B, Shih G, et al. ChatGPT and other large language models are double-edged swords. *Radiology*. 2023; 307(2): e230163.
11. Siddique S, Chow JCL. Machine learning in healthcare communication. *Encyclopedia*. 2021; 1(1): 220-39.
12. van Dis EAM, Bollen J, Zuidema W, van Rooij R, Bockting CL. ChatGPT: five priorities for research. *Nature*. 2023; 614(7947): 224-6.
13. Momenaei B, Wakabayashi T, Shahlaee A, Durrani AF, Pandit SA, Wang K, et al. Appropriateness and readability of ChatGPT-4-generated responses for surgical treatment of retinal diseases. *Ophthalmol Retina*. 2023; 7(10): 862-8.
14. Norman G. Likert scales, levels of measurement and the “laws” of statistics. *Adv Health Sci Educ Theory Pract*. 2010; 15(5): 625-32.
15. Readability is an essential content marketing tool. Readable. <https://readable.com/readability/#goodscore>.

16. Patel AJ, Kloosterboer A, Yannuzzi NA, Venkateswaran N, Sridhar J. Evaluation of the content, quality, and readability of patient accessible online resources regarding cataracts. *Semin Ophthalmol*. 2021; 36(5-6): 384-91.
17. Basch CH, Mohlman J, Hillyer GC, Garcia P. Public health communication in time of crisis: readability of on-line COVID-19 information. *Disaster Med Public Health Prep*. 2020; 14(5): 635-7.
18. Robinson E, McMenemy D. 'To be understood as to understand': a readability analysis of public library acceptable use policies. *J Librariansh Inf Sci*. 2020; 52(39): 713-25.
19. Lim ZW, Pushpanathan K, Yew SME, Lai Y, Sun CH, Lam JSH, et al. Benchmarking large language models' performances for myopia care: a comparative analysis of ChatGPT-3.5, ChatGPT-4.0, and Google Bard. *EBioMedicine*. 2023; 95: 104770.
20. Rasmussen MLR, Larsen AC, Subhi Y, Potapenko I. Artificial intelligence-based ChatGPT chatbot responses for patient and parent questions on vernal keratoconjunctivitis. *Graefes Arch Clin Exp Ophthalmol*. 2023; 261(10): 3041-3.
21. Mihalache A, Grad J, Patil NS, Huang RS, Popovic MM, Mallipatna A, et al. Google Gemini and Bard artificial intelligence chatbot performance in ophthalmology knowledge assessment. *Eye (Lond)*. 2024; 38(13): 2530-5.
22. Antaki F, Chopra R, Keane PA. Vision-language models for feature detection of macular diseases on optical coherence tomography. *JAMA Ophthalmol*. 2024; 142(6): 573-6.
23. Cheong KX, Zhang C, Tan TE, Fenner BJ, Wong WM, Teo KY, et al. Comparing generative and retrieval-based chatbots in answering patient questions regarding age-related macular degeneration and diabetic retinopathy. *Br J Ophthalmol*. 2024; 108(10): 1443-9.
24. Wang S, Wang Y, Jiang L, Chang Y, Zhang S, Zhao K, et al. Assessing the clinical support capabilities of ChatGPT 4o and ChatGPT 4o mini in managing lumbar disc herniation. *Eur J Med Res*. 2025; 30(1): 45.
25. Weiss BD. Health literacy and patient safety: help patients understand. Manual for clinicians. 2nd ed. Chicago, IL: American Medical Association Foundation and American Medical Association; 2007.
26. Redick DW, Hwang JC, Kloosterboer A, Yannuzzi NA, Patel NA, Kuriyan AE, et al. Content, readability, and accountability of freely available online information for patients regarding epiretinal membranes. *Semin Ophthalmol*. 2022; 37(1): 67-70.
27. Ozturk Y, Gungor CB. Assessing the quality, reliability, and readability of online information on rhegmatogenous retinal detachment quality information retinal detachment. *Research Square (Preprint)*. 2022.
28. Hillmann HAK, Angelini E, Karfoul N, Feickert S, Mueller-Leisse J, Duncker D. Accuracy and comprehensibility of chat-based artificial intelligence for patient information on atrial fibrillation and cardiac implantable electronic devices. *Europace*. 2023; 26(1): eua369.
29. Mu X, Lim B, Seth I, Xie Y, Cevik J, Sofiadellis F, et al. Comparison of large language models in management advice for melanoma: Google's AI BARD, BingAI and ChatGPT. *Skin Health Dis*. 2023; 4(1): e313.
30. Aggarwal A, Tam CC, Wu D, Li X, Qiao S. Artificial intelligence-based chatbots for promoting health behavioral changes: systematic review. *J Med Internet Res*. 2023; 25: e40789.
31. Abd-Alrazaq A, AlSaad R, Alhuwail D, Ahmed A, Healy PM, Latifi S, et al. Large language models in medical education: opportunities, challenges, and future directions. *JMIR Med Educ*. 2023; 9: e48291.
32. Li Z, Li F, Fu Q, Wang X, Liu H, Zhao Y, et al. Large language models and medical education: a paradigm shift in educator roles. *Smart Learn*. 2024; 11(1): 26.
33. Moldt JA, Festl-Wietek T, Madany Mamlouk A, Nieselt K, Fuhl W, Herrmann-Werner A. Chatbots for future docs: exploring medical students' attitudes and knowledge towards artificial intelligence and medical chatbots. *Med Educ Online*. 2023; 28(1): 2182659.
34. Wah JNK. Revolutionizing e-health: the transformative role of AI-powered hybrid chatbots in healthcare solutions. *Front Public Health*. 2025; 13: 1530799.
35. Amin KS, Mayes LC, Khosla P, Doshi RH. Assessing the efficacy of large language models in health literacy: a comprehensive cross-sectional study. *Yale J Biol Med*. 2024; 97(1): 17-27.

Asymptomatic Anterior Tarsal Tunnel Syndrome: An Underdiagnosed Entity

✉ Filiz Azman-İste¹, ✉ Berfe Simya Özbek²

¹Clinic of Neurology, Division of Clinical Neurophysiology, University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, İstanbul, Türkiye

²Clinic of Neurology, University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital, İstanbul, Türkiye

Abstract

BACKGROUND/AIMS: Anterior tarsal tunnel syndrome (ATTS) is an uncommon entrapment neuropathy of the deep peroneal nerve that is frequently overlooked in clinical practice. Prayer-related and ground-sitting postures, which are culturally common in Türkiye, may predispose individuals to this condition. This study aims to evaluate the prevalence, clinical characteristics, and electrophysiological features of asymptomatic ATTS identified incidentally during routine electromyography (EMG) examinations.

MATERIALS AND METHODS: The records of 3,610 patients who underwent electrophysiological evaluation between January 2025 and February 2026 were retrospectively reviewed. Patients with electrophysiological evidence of distal deep peroneal nerve involvement without ATTS-related symptoms were included. L5 radiculopathy, proximal peroneal neuropathy, and polyneuropathy with motor fiber involvement were excluded.

RESULTS: Asymptomatic ATTS was identified in 21 patients (0.58%). The mean age was 56.1 ± 11.9 years; 11 were female and 10 were male. All patients reported prolonged plantar-flexion postures, mainly prayer-related sitting habits (namaz), with a mean duration of 15 years. Ten patients had diabetes mellitus. On the affected side, mean extensor digitorum brevis (EDB) distal motor latency was 5.19 ± 1.57 ms and mean compound muscle action potential amplitude was 1.04 ± 0.52 mV, compared to 4.47 ± 0.8 ms and 3.08 ± 1.52 mV on the contralateral side. Tibialis anterior (TA) motor responses were normal bilaterally. Needle EMG demonstrated neurogenic changes confined to the EDB muscle in all patients.

CONCLUSION: Asymptomatic ATTS may be encountered more frequently in EMG practice than is generally appreciated. Prayer-related and ground-sitting postures may predispose individuals to ATTS. Systematic electrophysiological evaluation, including TA recording and needle EMG, may reduce misdiagnosis of polyneuropathy or lumbosacral radiculopathy in daily practice.

Keywords: Anterior tarsal tunnel syndrome, deep peroneal nerve, electromyography, entrapment neuropathy, prayer posture

INTRODUCTION

Anterior tarsal tunnel syndrome (ATTS) is an uncommon and often underdiagnosed entrapment neuropathy of the deep peroneal nerve at the ankle level. The anterior tarsal tunnel is a flattened fibro-osseous space located between the tendons of the extensor hallucis longus and extensor digitorum longus muscles, bounded superficially by the inferior

extensor retinaculum and deeply by the fascia overlying the talus and navicular bones.^{1,2} The tunnel is narrow proximally and widens distally; its contents include the deep peroneal nerve and the dorsalis pedis artery, both of which are relatively unprotected given the superficial location of this space. Within or just distal to the tunnel, the deep peroneal nerve divides into two terminal branches: the lateral branch, which provides motor innervation to the extensor digitorum brevis (EDB) muscle, and

To cite this article: Azman-İste F, Özbek BS. Asymptomatic anterior tarsal tunnel syndrome: an underdiagnosed entity. Cyprus J Med Sci. 2026;11(5):378-383

ORCID IDs of the authors: F.A.İ. 0000-0002-4032-1613; B.S.Ö. 0009-0000-0010-4430.



Corresponding author: Filiz Azman-İste
E-mail: filizazman@hotmail.com
ORCID ID: orcid.org/0000-0002-4032-1613

Received: 23.06.2026
Accepted: 24.07.2026
Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.
This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

the medial branch, which supplies sensation to the first interdigital web space.¹ Compression of the deep peroneal nerve within this tunnel results in ATTS, which is typically associated with dorsal foot pain, paresthesia in the first interdigital web space, and, in some cases, weakness or atrophy of the EDB muscle.^{3,4}

ATTS varies widely in clinical severity and does not always present with all typical findings. Symptoms may be mild or non-specific, and nerve involvement can remain entirely asymptomatic in some patients, making the diagnosis easy to overlook.^{5,6} In these cases, electrophysiological studies, particularly those evaluating the distal deep peroneal nerve, can play a key role in detecting this abnormality. Nevertheless, asymptomatic or subclinical presentations of ATTS are rarely reported, and the condition is mostly discussed in symptomatic patients in the literature.⁷⁻¹¹

In our clinical practice, we observed that ATTS may be encountered more frequently than expected, which may be related to cultural habits such as sitting on the floor and postures adopted during prayer (namaz), both of which involve sustained ankle stretching. Notably, these patients did not report symptoms suggestive of ATTS. Based on this observation, the present study aims to evaluate patients in whom electrophysiological findings consistent with ATTS were incidentally identified during routine electromyography (EMG) examinations. We aimed to examine their clinical and electrophysiological characteristics in detail and to explore potential etiological factors, including daily habits and related risk factors.

MATERIALS AND METHODS

Study Design

The records of 3,610 patients who underwent electrophysiological evaluation at the electroneuromyography (ENMG) laboratory at University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital between January 2025 and February 2026 were retrospectively reviewed. All studies were performed by a clinical neurophysiologist. Patients referred for various clinical indications were screened, and those without symptoms suggestive of ATTS but with electrophysiological evidence of distal deep peroneal nerve involvement were included in the study. Asymptomatic ATTS was defined as electrophysiological evidence of distal deep peroneal nerve entrapment in the absence of ATTS-related symptoms.

Patients were included if they were 18 years of age or older; had abnormal compound muscle action potential (CMAP) recordings from the EDB muscle with normal CMAP recordings from the tibialis anterior (TA) muscle; had neurogenic changes on concentric needle EMG confined to the EDB; had normal contralateral peroneal nerve CMAP recordings from both the EDB and TA muscles; and had no ATTS-related clinical complaints.

Patients were excluded if they were younger than 18 years of age, had clinical or electrophysiological evidence of L5 radiculopathy, proximal peroneal neuropathy, or polyneuropathy involving motor fibers, had a history of ankle trauma or surgery, presented with ATTS-related symptoms, or had incomplete electrophysiological data.

Electrophysiological Studies

During the electrophysiologic studies, skin temperature was maintained above 32 °C. Nerve conduction studies (NCS) were performed using

standard techniques with surface electrodes. Bandpass filters were set at 3 Hz-10 kHz for motor conduction studies and at 20 Hz-3 kHz for sensory conduction studies. Sweep speeds were set at 5 ms/division and 2 ms/division for motor and sensory conduction studies, respectively. Sensitivity settings of 10-20 µV/division and 5 mV/division were used for sensory and motor studies, respectively.

Bilateral sural sensory NCS were performed in all patients. In patients where technically feasible, bilateral deep peroneal sensory NCS were performed using the antidromic method, with recording electrodes placed over the first interdigital space, as previously described.^{12,13} For deep peroneal sensory NCS, the normative values reported by Lee et al.¹² were used [onset latency: 2.9 ± 0.4 ms and sensory nerve action potential (SNAP) amplitude: 3.4 ± 1.2 µV]. Motor NCS included peroneal nerve recordings from the EDB and TA muscles and tibial nerve recordings from the abductor hallucis muscle. Normative reference values for peroneal motor NCS recorded from the EDB and TA muscles, as well as for all other routine motor and sensory NCS, were based on the reference values described by Preston and Shapiro,¹⁴ which are routinely used in our electrophysiology laboratory. For peroneal motor recordings, the upper limits of normal distal motor latency (DML) were 6.5 ms for the EDB and 6.7 ms for the TA; the lower limits of normal CMAP amplitude were 2.0 mV for the EDB and 3.0 mV for the TA. All findings were interpreted in comparison with the contralateral side. Needle EMG was performed on all patients, including routine examination of the EDB and TA muscles on the affected side. Additional muscles were examined when clinically indicated.

Demographic and Clinical Data

In all patients diagnosed with asymptomatic ATTS, demographic and clinical data were collected, including age, sex, and history of chronic diseases. Potential risk factors were systematically assessed, with particular attention to daily habits that may lead to sustained compression of, or repetitive microtrauma to the extensor retinaculum at the ankle. Specifically, sitting habits involving prolonged ankle flexion, postures adopted during Islamic prayer (namaz), and the durations of these habits were recorded. Occupational characteristics were also documented. These factors were assessed to explore their potential contribution to the development of distal deep peroneal nerve entrapment. In addition to postural habits, potential risk factors that may predispose to nerve compression, including diabetes mellitus, vasculitis, and other relevant comorbidities and their duration were recorded for each patient.

Ethical Approval

Ethical approval was obtained from the University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital Clinical Research Ethical Committee (approval number: 237, date: 22.04.2026). Patient consent was not obtained because the study was retrospective. The Declaration of Helsinki and good clinical practice procedures were followed in the study.

Statistical Analysis

Descriptive statistics were used to analyze the data. Continuous variables were expressed as mean $\pm$ standard deviation or median values, depending on data distribution, while categorical variables were presented as frequencies and percentages.

RESULTS

Asymptomatic ATTS was identified in 21 of 3,610 patients (0.58%). The demographic characteristics and risk factors of all patients are summarized in Table 1. Eleven patients were female and 10 were male, and the mean age was 56.1 ± 11.9 years (range 27-77 years). Fifteen patients had right-sided ATTS involvement, and six had left-sided ATTS involvement. None of the patients reported symptoms consistent with ATTS despite the presence of clear electrophysiological abnormalities. All patients reported postural habits involving prolonged plantar flexion of the ankle, including prayer-related sitting positions in 18 patients, a combination of prayer-related postures and general sitting habits in two patients, and occupational exposure (tailoring and regular pedal use) in one patient. The mean duration of these habits was 15 years (range: 8-30 years). Additionally, 10 patients had a history of diabetes mellitus.

Electrophysiological evaluation revealed notably differences between the affected and unaffected sides, as detailed in Table 2. In three patients, no motor response could be obtained from peroneal nerve stimulation recorded from the EDB muscle; however, normal responses were obtained from the contralateral side (mean DML: 4.2 ms; mean CMAP: 3.9 mV). Peroneal nerve motor responses recorded from the TA muscle were normal on both the affected and contralateral sides (DMLs 2.85 ms and 3.5 ms, and CMAP amplitudes 4.2 mV and 4.3 mV, respectively). In the remaining 18 patients in whom an EDB-recorded response was obtainable at the ATTS side, the mean DML on the affected side was 5.19 ± 1.57 ms with a mean CMAP amplitude of 1.04 ± 0.52 mV, compared to a mean DML of 4.47 ± 0.8 ms and a mean amplitude of 3.08 ± 1.52 mV on the contralateral side, respectively. TA-recorded motor responses showed no side-to-side difference: on the affected side, mean DML was 3.1 ± 0.5 ms with a mean amplitude of 4.08 ± 2 mV, while on the contralateral side mean DML was 2.9 ± 0.5 ms and mean amplitude was 4.2 ± 1.7 mV (Table 2).

Deep peroneal sensory nerve conduction was assessed in 11 patients, of whom eight had bilaterally absent sensory responses. Among the remaining three patients, two had no response on the affected side, while normal responses were recorded on the contralateral side. In one patient, a reduced SNAP amplitude was noted on the affected side ($4 \mu\text{V}$, latency: 2.8 ms) compared with the contralateral side ($10 \mu\text{V}$, latency: 2.4 ms) (Table 2).

Needle EMG of the EDB muscle demonstrated neurogenic changes in all patients. Rare denervation potentials were observed in the EDB muscle in one patient. During mild voluntary contraction, chronic neurogenic

changes in the EDB muscle were graded as moderate-to-severe in 15 patients and as mild in six patients.

DISCUSSION

In this study, we present a series of patients who had electrophysiological findings consistent with ATTS despite the absence of ATTS-related clinical symptoms. While ATTS is generally described as a symptomatic condition characterized by dorsal foot pain and sensory changes in the first web space, our results support the presence of a subclinical or asymptomatic form that may be more common than previously recognized. The prevalence observed in our study (0.58%) is comparable to that reported by Logullo et al.,⁵ who identified ATTS in 85 of 19,582 patients (0.43%) examined over a ten-year period in a clinical neurophysiology laboratory. In their series, the majority of patients were asymptomatic or had only mild, non-specific complaints, such as a dull ache on the dorsum of the foot. These findings suggest that asymptomatic ATTS may be regularly encountered in ENMG practice when specifically sought,

Table 2. Electrophysiological findings of patients with asymptomatic anterior tarsal tunnel syndrome

Parameter	Affected side	Contralateral side
Motor nerve conduction studies-peroneal nerve		
NR on EDB at affected side (n=3, 15%)		
EDB recording		
DML, ms	NR	4.2
CMAP amplitude, mV	NR	3.9
TA recording		
DML, ms	2.85	3.5
CMAP amplitude, mV	4.2	4.3
EDB response present at affected side (n=18, 85%)		
EDB recording		
DML, mean $\pm$ SD, ms	5.19 $\pm$ 1.57	4.47 $\pm$ 0.80
CMAP amplitude, mean $\pm$ SD, mV	1.04 $\pm$ 0.52	3.08 $\pm$ 1.52
TA recording		
DML, mean $\pm$ SD, ms	3.1 $\pm$ 0.5	2.9 $\pm$ 0.5
CMAP amplitude, mean $\pm$ SD, mV	4.08 $\pm$ 2.0	4.2 $\pm$ 1.7
Sensory nerve conduction studies-deep peroneal nerve		
Patients assessed	n=11	
No response bilaterally	n=8 (72.7%)	
NR affected/normal contralateral (n=2)	NR	1. patient: 9.9 μ V/2.71 ms 2. patient: 7.5 μ V/3.02 ms
SNAP amplitude/onset latency at affected & contralateral (n=1)	4 μ V/2.8 ms	10 μ V/2.4 ms
Needle EMG-EDB muscle (n=21)		
Spontaneous denervation potentials	1 (4.8%)	
Neurogenic changes	n=21 (100%)	
Moderate-to-severe	15 (71.4%)	
Mild	6 (28.6%)	
Values are mean $\pm$ SD unless otherwise stated.		
DML: Distal motor latency, CMAP: Compound muscle action potential, SNAP: Sensory nerve action potential, EDB: Extensor digitorum brevis, TA: Tibialis anterior, NR: No response, NCS: Nerve conduction study, EMG: Electromyography, SD: Standard deviation.		

Table 1. Demographic characteristics and risk factors of patients with asymptomatic anterior tarsal tunnel syndrome

Demographics	
Age, mean $\pm$ SD (range), years	56.1 ± 11.9 (27-77)
Sex, female/male	11/10
Side affected, right/left	15/6
Diabetes mellitus, n (%)	10 (47.6%)
Postural risk factors	
Prayer-related sitting only, n (%)	18 (85.7%)
Prayer-related + general sitting, n (%)	2 (9.5%)
Occupational exposure, n (%)	1 (4.8%)
Duration of postural habits, mean (range), years	15 (8-30)
SD: Standard deviation.	

and that its real frequency may be underestimated due to a lack of awareness. However, both studies were conducted in EMG laboratory settings, and the figures obtained therefore reflect referral-based populations rather than the true prevalence in the general population.

Entrapment neuropathies are generally considered clinical disorders, with electrodiagnostic studies serving as complementary tools to confirm the diagnosis, localize the lesion, assess its severity, and exclude alternative diagnoses.^{15,16} Accordingly, electrodiagnostic evaluation has become an integral component in the assessment of suspected entrapment neuropathies, particularly when the clinical presentation is atypical or the diagnosis is uncertain. Nevertheless, our findings suggest that ATTS may also exist in a subclinical form, in which characteristic electrophysiological abnormalities are present despite the absence of typical symptoms. In such cases, electrodiagnostic studies may identify distal deep peroneal nerve entrapment before overt clinical manifestations appear.

The most notable finding in our cohort was the strong association between asymptomatic ATTS and postures involving chronic stretching and compression of the deep peroneal nerve. Eighteen patients had prayer-related sitting habits (namaz), two had both prayer and general ground-sitting habits, and one had occupational exposure involving regular pedal use and tailoring. The mean duration of these habits was 15 years. Previous studies from Türkiye have also linked ATTS to prayer-related and ground-sitting postures.^{8,17} Chronic stretching of the deep peroneal nerve during kneeling and prolonged sitting positions in namaz, together with body weight-related compression at the anterior tarsal tunnel, have been proposed as the underlying mechanism.⁸ Sociocultural sitting habits and repetitive mechanical exposure may, therefore, play an important role in the development of ATTS. Our series is the largest prayer-related ATTS cohort reported to date, and the first to demonstrate this association exclusively in asymptomatic individuals. Beyond prayer-related posture, deep peroneal nerve compression has also been reported in other repetitive occupational and mechanical settings, including toddy tappers in South India and dancers wearing tight footwear, suggesting that diverse cultural and occupational exposures may converge on a shared compressive mechanism at the anterior tarsal tunnel.^{6,18}

The conventional Islamic prayer sitting posture is characterized by the left foot positioned beneath the body in plantar flexion and the right foot upright with the toes flexed and the ankle in slight dorsiflexion. In this position, the right foot appears to be subjected to greater mechanical stress, which is consistent with the predominance of right-sided involvement observed in our cohort (71%). However, when patients were asked to describe their habitual sitting posture during prayer, considerable individual variation was observed. These variations may also explain the occurrence of left-sided involvement in some patients. Indeed, previous reports have described right-sided, left-sided, as well as bilateral ATTS.^{8,17} In addition, patients with habitual floor-sitting practices reported adopting sitting positions in which the right knee was flexed and the right foot was positioned beneath the contralateral leg with the ankle in dorsiflexion, potentially placing greater mechanical stress on the right foot and the anterior tarsal tunnel. Our observations suggest that individual variations in prayer and floor-sitting positions may generate varying degrees and distributions of mechanical stress on the deep peroneal nerve and further contribute to the observed predominance of right-sided involvement among patients in our cohort.

The anterior tarsal tunnel is a narrow anatomical region where the deep peroneal nerve is particularly susceptible to external pressure.^{1,2} In our cohort, all patients reported long-standing postural habits characterized by ankle plantar flexion. Repetitive stretching and compression of the nerve during these postures may contribute to gradual nerve injury. Unlike acute compression injuries, chronic low-grade mechanical stress may develop slowly and remain clinically silent for years. This may explain why clear electrophysiological abnormalities were detected despite the absence of symptoms in our patients. Similar mechanisms have been suggested in previous reports of postural and occupational forms of ATTS.^{6,19} Although this remains speculative, all patients in our cohort reported similar long-standing postural habits, which suggests that chronic mechanical stress may contribute to asymptomatic involvement of the distal deep peroneal nerve.

The absence of symptoms despite clear electrophysiological abnormalities may partly reflect the limited functional role of the EDB muscle, whose function is largely compensated for by the extensor digitorum longus and extensor hallucis longus.^{3,6} Additionally, ten of our patients had diabetes mellitus, a condition known to increase susceptibility to entrapment neuropathies through axonal vulnerability and impaired nerve repair mechanisms. While diabetes alone was not sufficient to cause symptoms in our cohort, its coexistence with chronic postural compression may lower the threshold for nerve injury and contribute to more severe electrophysiological findings.

The electrophysiological findings were consistent across all patients. On the affected side, peroneal motor responses recorded from the EDB showed prolonged distal latency and reduced amplitude, while contralateral EDB responses were normal. In three patients, no response was obtained from the affected side. TA motor responses were normal and bilaterally symmetric in all patients. Needle EMG demonstrated chronic neurogenic changes confined to the EDB muscle, with normal findings in all other examined muscles, thereby excluding lumbosacral radiculopathy and polyneuropathy, and confirming a focal axonal lesion at the anterior tarsal tunnel. This distribution also argued against age-related EDB atrophy or technical artifact as alternative explanations.^{3,5} Sensory deep peroneal nerve conduction was evaluated in only 11 patients because of the well-known technical challenges in recording sensory responses from this nerve, particularly in older individuals. In this context, motor conduction studies provide more reliable and consistent results in the diagnosis of ATTS. Among the 11 patients studied, responses were absent bilaterally in 8, absent on the affected side only in 2, and reduced in amplitude in 1. The asymmetric findings observed in three patients provided additional support for a focal lesion at the anterior tarsal tunnel.³

ATTS is frequently underdiagnosed in routine ENMG practice, as isolated EDB abnormalities may be attributed to polyneuropathy, lumbosacral radiculopathy, or age-related changes.¹⁹ In this context, a clinically important diagnostic pitfall was previously raised by Al-Hashel et al.²⁰ in response to McNeish et al.,²¹ who proposed that a fibular motor CMAP amplitude below 3.6 mV recorded from the EDB may predict L5 radiculopathy in patients with radiating limb pain. Al-Hashel et al.²⁰ pointed out that this threshold may be misleading in Middle Eastern and Asian populations, where habitual floor-sitting and kneeling during Islamic prayer involve sustained plantar flexion of the ankle, potentially compressing the EDB muscle and the lateral branch of the deep peroneal nerve.²² Similarly, Kishimoto et al.²² reported that bilateral EDB atrophy is significantly associated with distal symmetric polyneuropathy in

Japanese men, but only after excluding individuals with a seiza habit (the traditional Japanese style of kneeling and sitting over one's heels) as this posture independently causes EDB atrophy through a mechanism similar to ATTS. Our findings are consistent with these observations. All 21 patients in our cohort had EDB CMAP amplitudes below 3.6 mV, yet none had evidence of L5 radiculopathy or polyneuropathy. Therefore, in populations where prayer-related or traditional floor-sitting postures are common, isolated EDB abnormalities should be interpreted with caution, and focal entrapment at the anterior tarsal tunnel should be considered. Systematic TA recording, needle EMG evaluation, and, when available, deep peroneal sensory nerve studies may help improve diagnostic accuracy and prevent misattribution of low EDB CMAP amplitudes to proximal lesions.^{5,6}

Additionally, one patient reported long-term tailoring work requiring repetitive use of a sewing machine pedal. Although tailoring has not been specifically reported as a risk factor for ATTS, repetitive ankle movements and chronic microtrauma associated with pedal use may contribute to deep peroneal nerve compression, consistent with the proposed pathophysiology of ATTS.²³

Study Limitations

This study has several limitations. First, its retrospective design may have introduced selection bias and limited the availability of clinical data. Second, because the cohort was derived from an EMG laboratory population, the reported prevalence reflects a referral-based setting and may not represent the general population. Third, deep peroneal sensory NCS could only be performed in a subset of patients, primarily because of technical difficulties in older patients, which may have led to under-recognition of sensory involvement. Fourth, imaging studies were not routinely available; therefore, structural causes of nerve compression such as osteophytes, ganglia, or soft tissue lesions could not be completely excluded. Finally, no longitudinal follow-up was performed, and the long-term clinical course of asymptomatic ATTS remains uncertain.

CONCLUSION

Asymptomatic ATTS may be encountered more frequently in EMG practice than is generally expected. In societies where prayer-related and ground-sitting postures are culturally common, these habits represent important risk factors for ATTS. Routine inquiry into such postural habits, combined with bilateral systematic electrophysiological evaluation, including TA recordings, needle EMG, and deep peroneal sensory nerve studies, where feasible, may improve diagnostic accuracy and reduce the risk of misdiagnosis of polyneuropathy or lumbosacral radiculopathy.

MAIN POINTS

- Asymptomatic anterior tarsal tunnel syndrome is more common than previously recognized, with a prevalence of 0.58% in an electromyography laboratory setting.
- Prayer-related and ground-sitting postures are the most frequent identifiable risk factors.
- Electrophysiological abnormalities were confined to the extensor digitorum brevis (EDB) muscle, confirming a focal lesion at the anterior tarsal tunnel.

- Low EDB compound muscle action potential amplitudes in populations with habitual floor-sitting should not be misattributed to L5 radiculopathy or polyneuropathy.
- Systematic electrophysiological evaluation and routine inquiry into postural habits may reduce misdiagnosis in daily practice.

ETHICS

Ethics Committee Approval: Ethical approval was obtained from the University of Health Sciences Türkiye, Başakşehir Çam and Sakura City Hospital Clinical Research Ethical Committee (approval number: 237, date: 22.04.2026).

Informed Consent: Patient consent was not obtained because the study was retrospective.

Footnotes

Authorship Contributions

Surgical and Medical Practices: F.A.-İ., B.S.Ö., Concept: F.A.-İ., Design: F.A.-İ., B.S.Ö., Data Collection and/or Processing: F.A.-İ., B.S.Ö., Analysis and/or Interpretation: F.A.-İ., B.S.Ö., Literature Search: B.S.Ö., Writing: F.A.-İ.

DISCLOSURES

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

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

Declaration on the Use of Artificial Intelligence (AI): Artificial intelligence has been used to assist with "language correction".

REFERENCES

1. Aktan İkiz ZA, Ucerler H, Uygur M. Dimensions of the anterior tarsal tunnel and features of the deep peroneal nerve in relation to clinical application. *Surg Radiol Anat.* 2007; 29(7): 527-30.
2. Banu J, Dhakshnamoorthy N, Sakthivel S. Unravelling the anatomy of the anterior tarsal tunnel and its clinical implications. *Cureus.* 2024; 16(7): e64282.
3. Oh SJ. Neuropathies of the foot. *Clin Neurophysiol.* 2007; 118(5): 954-80.
4. DiDomenico LA, Masternick EB. Anterior tarsal tunnel syndrome. *Clin Podiatr Med Surg.* 2006; 23(3): 611-20.
5. Logullo F, Ganino C, Lupidi F, Perozzi C, Di Bella P, Provinciali L. Anterior tarsal tunnel syndrome: a misunderstood and a misleading entrapment neuropathy. *Neurol Sci.* 2014; 35(5): 773-5.
6. Panwar A, Madhavarao V, Mohammed O, Valupadas C. High prevalence of asymptomatic anterior tarsal tunnel syndrome in toddy tappers of south india: a case series of 21 participants. *J Neurosci Rural Pract.* 2019; 10(3): 506-10.
7. Bianchi S, Droz L, Le Corroller T, Delmi M. Partial anterior tunnel syndrome: a retrospective analysis of ultrasound findings in four surgically proven cases. *Skeletal Radiol.* 2019; 48(5): 807-12.
8. Secil Y, Tokucoglu F, Beckmann Y, Arici Ş, Eryasar G. Anterior tarsal tunnel syndrome: electrophysiological and clinical evaluation of five cases. *J Neurol Sci [Turk].* 2012; 29(4): 819-25.
9. Krause KH, Witt T, Ross A. The anterior tarsal tunnel syndrome. *J Neurol.* 1977; 217(1): 67-74.

10. Sillat T, Pivec C, Bernathova M, Moritz T, Bodner G. Unusual cause of anterior tarsal tunnel syndrome: ultrasound findings. *J Ultrasound Med*. 2017; 36(4): 837-9.
11. Yassin M, Garti A, Weissbrot M, Heller E, Robinson D. Treatment of anterior tarsal tunnel syndrome through an endoscopic or open technique. *Foot (Edinb)*. 2015; 25(3): 148-51.
12. Lee HJ, Bach JR, DeLisa JA. Deep peroneal sensory nerve. Standardization in nerve conduction study. *Am J Phys Med Rehabil*. 1990; 69(4): 202-4.
13. Lo YL, Leoh TH, Dan YF, Tan YE, Nurjannah S, Fook-Chong S. An electrophysiological study of the deep peroneal sensory nerve. *Eur Neurol*. 2003; 50(4): 244-7.
14. Preston DC, Shapiro BE. Electromyography and neuromuscular disorders: clinical-electrodiagnostic-ultrasound correlations, fourth edition. *J Clin Neurophysiol*. 2021; 38(4): e19.
15. Temuçin ÇM, Yıldız FG. Electrophysiological examinations in entrapment neuropathies. *TOTBİD Dergisi*. 2015; 14: 481-7.
16. Pugdahl K, Tankisi H, Fuglsang-Frederiksen A. Electrodiagnostic testing of entrapment neuropathies: a review of existing guidelines. *J Clin Neurophysiol*. 2020; 37(4): 299-305.
17. Akyüz G, Us O, Türan B, Kayhan O, Canbulat N, Yilmaz IT. Anterior tarsal tunnel syndrome. *Electromyogr Clin Neurophysiol*. 2000; 40(2): 123-8.
18. Kennedy JG, Baxter DE. Nerve disorders in dancers. *Clin Sports Med*. 2008; 27(2): 329-34.
19. Yan Y, Wang D, Kalia V, Garvin G. MRI findings of bilateral anterior tarsal tunnel syndrome - a case report. *J Musculoskelet Neuronal Interact*. 2024; 24(4): 433-8.
20. Al-Hashel JY, Rousseff RT, Khuraibet AJ, Rashad HM. Isolated asymptomatic atrophy of the extensor digitorum brevis muscle. *Muscle Nerve*. 2020; 61(1): E10-1.
21. McNeish B, Hearn S, Craig A, Laidlaw A, Ziadeh M, Richardson JK. Motor amplitudes may predict electromyography-confirmed radiculopathy in patients referred for radiating limb pain. *Muscle Nerve*. 2019; 59(5): 561-6.
22. Kishimoto S, Sasaki H, Kurisu S, Ogawa K, Matsuno S, Furuta H, et al. Bilateral atrophy of the extensor digitorum brevis muscle might be a useful sign for diagnosing diabetic polyneuropathy in Japanese men who do not sit in the traditional "seiza" style. *J Diabetes Investig*. 2021; 12(3): 398-408.
23. Gonzalez J, Bergman R. Anterior tarsal tunnel syndrome. 2026 May 24. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026 Jan-.

Temporomandibular Joint Arthrocentesis: A Bibliometric Analysis From 1989 to 2024

✉ Mustafa İşleyen, ✉ Asenur Aydemir

Department of Oral and Maxillofacial Surgery, Mehmet Akif Ersoy University Faculty of Dentistry, Burdur, Türkiye

Abstract

BACKGROUND/AIMS: Arthrocentesis of the temporomandibular joint (TMJ), introduced in the early 1990s, has become a cornerstone of minimally invasive management for TMJ internal derangements; however, the scientific structure and evolution of research in this field have not been comprehensively evaluated. This study aimed to analyze the scientific development, research trends, and collaborative patterns in TMJ arthrocentesis using bibliometric methods.

MATERIALS AND METHODS: A total of 581 publications (1989-2024) were retrieved from the Web of Science Core Collection. Bibliometric analyses were performed using biblioshiny to evaluate publication trends, citation performance, leading authors, institutions, and journals. VOSviewer was employed to visualize author, institutional, and country collaboration networks, as well as article co-citation structures and keyword co-occurrence patterns, allowing the identification of influential research clusters and thematic evolution.

RESULTS: The United States and the Karolinska Institute demonstrated the highest contributions at the national and institutional levels, respectively. Manfredini D. and Guarda-Nardini L. were the most prolific authors, with the Journal of Oral and Maxillofacial Surgery as the primary publication venue. Common keywords included “arthrocentesis”, “temporomandibular joint”, and “temporomandibular joint disorders”. Bibliometric mapping revealed a temporal shift in thematic trends: whereas initial terms focused on mechanical concepts, regenerative agents such as “platelet-rich plasma” and “hyaluronic acid” showed increased co-occurrence density after 2020.

CONCLUSION: Current trends indicate that arthrocentesis is evolving from a traditional irrigation-based technique into a regenerative therapeutic modality. This bibliometric analysis provides a comprehensive overview of influential contributors, emerging topics, and evolving themes, offering guidance for clinicians and researchers seeking to advance the field.

Keywords: Arthrocentesis, bibliometrics, regenerative medicine, temporomandibular joint disorders

INTRODUCTION

Temporomandibular disorders (TMDs) are musculoskeletal conditions characterized by pain and dysfunction of the temporomandibular joint (TMJ). They represent the third most common oral and maxillofacial disorders after dental caries and periodontal disease. TMDs, which are more frequently observed in young adults and women, may present as acute or chronic conditions and are commonly associated with symptoms such as pain, limited mouth opening, and joint noises during

mandibular movement. These clinical manifestations can significantly impair patients' quality of life.^{1,2}

Management strategies for TMDs encompass a broad spectrum of treatment modalities, ranging from conservative, non-surgical approaches to minimally invasive interventions, and open joint surgery. Conservative therapies, such as patient education, physiotherapy, occlusal splint therapy, and pharmacological agents, including non-steroidal anti-inflammatory drugs and muscle relaxants, are generally

To cite this article: İşleyen M, Aydemir A. Temporomandibular joint arthrocentesis: a bibliometric analysis from 1989 to 2024. Cyprus J Med Sci. 2026;11(5):384-392

ORCID IDs of the authors: M.İ. 0000-0003-3326-0047; A.A. 0009-0007-8667-3933.



Corresponding author: Mustafa İşleyen

E-mail: misleyen42@gmail.com

ORCID ID: orcid.org/0000-0003-3326-0047

Received: 13.03.2026

Accepted: 27.07.2026

Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.

This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

recommended as first-line treatment options. When conservative management fails to achieve sufficient pain relief or functional improvement, minimally invasive procedures, particularly TMJ arthrocentesis and arthroscopy, are commonly considered.³⁻⁵

TMJ arthrocentesis was first described by Nitzan et al.⁶ in 1991 as a minimally invasive technique involving lavage of the upper joint compartment using solutions such as Ringer's lactate or sterile saline. The procedure is typically performed under local anesthesia but may also be carried out under intravenous sedation or general anesthesia.⁶ Since its original description, several technical modifications have been introduced, and arthrocentesis procedures are broadly classified by the number of puncture sites used: single- or double-puncture arthrocentesis.⁷ Arthrocentesis has been widely applied in the management of TMDs, particularly for internal derangements (including disc displacement with or without reduction) and for degenerative joint diseases such as osteoarthritis. These conditions are frequently associated with masticatory pain at rest or during function, limitation of maximum mouth opening, and subsequent deterioration in oral function and quality of life.⁸

Bibliometrics is a quantitative research method used to assess scientific productivity, research impact, and the intellectual structure of a specific field through the analysis of academic publications.⁹ Bibliometric analyses commonly examine publication output, citation performance, authorship patterns, and institutional or country-level contributions, as well as thematic structures that reflect the evolution of research topics over time. When appropriately applied, bibliometric approaches provide objective insights into research trends, influential contributors, and emerging areas of scientific interest.¹⁰

Although bibliometric studies have been conducted in oral and maxillofacial surgery focusing on TMJ disorders, TMJ osteoarthritis, and highly cited publications, no bibliometric analysis has specifically addressed TMJ arthrocentesis.¹¹⁻¹⁵ Consequently, the scientific development, structure, and evolution of research in this area remain underexplored. Therefore, this study aimed to provide a comprehensive bibliometric analysis of TMJ arthrocentesis publications from 1989 and 2024 by evaluating publication trends, collaborative networks, influential journals and authors, and temporal changes in research themes, to inform and guide future investigations in this field.

MATERIALS AND METHODS

Data Source and Search Strategy

For our bibliometric analysis, we selected the Web of Science Core Collection (WoSCC), the globally recognized database platform provided by Clarivate Analytics, to obtain global academic data. The literature search was conducted on November 15, 2025, using the following strategy applied to the title, abstract, and author keywords fields: ("arthrocentesis" OR "lavage" OR "lysis" OR "irrigation" OR "joint washing" OR "lysis and lavage" OR "joint irrigation" OR "joint lavage therapy") AND ("TMJ" OR "temporomandibular" OR "TMJ" OR "craniomandibular joint") NOT (knee OR shoulder OR hip OR ankle OR elbow OR wrist). The retrieved records were refined according to predefined inclusion criteria. Publications indexed up to 2024 were included, while records from 2025 were excluded to avoid incomplete citation data. Publications from 2025 were excluded to ensure a complete citation window and to avoid bias related to incomplete indexing. Only documents classified as "article" or "review article" were considered. The analysis was restricted

to english-language publications indexed in the science citation index expanded and the emerging sources citation index and categorized under the Dentistry, Oral Surgery & Medicine category. Following the application of these criteria, a total of 581 publications were included in the final dataset (Figure 1). Bibliographic information, including publication year, title, authors, affiliations, countries, abstracts, keywords, and source journals, was exported in both TXT and BibTeX formats for subsequent bibliometric analysis. As no animals or humans were involved in the study, ethical approval was not required.

Eligibility Criteria and Language Selection

Only original research and review articles were included to ensure methodological consistency and reproducibility of bibliometric indicators. Document types with limited peer-review depth, including editorials, letters, conference proceedings, and case reports, were excluded from the analysis. The analysis was restricted to english-language publications. This language criterion was applied to maintain terminological consistency and enable reliable citation-based evaluation, as English is the predominant language of international scientific communication and indexing in bibliometric databases.

Bibliometric Analysis and Visualizations

Bibliometric analyses were performed using VOSviewer (version 1.6.20; Leiden University, Leiden, Netherlands) and the R package bibliometrix via its Biblioshiny interface (Vienna, Austria). VOSviewer was employed to construct and visualize collaboration networks among authors, institutions, and countries, as well as co-citation networks of references, authors, and journals. Keyword analysis was conducted using co-occurrence mapping to identify frequently associated terms and to reveal thematic structures and research hotspots within the field. Co-citation analysis was used to assess citation relationships: two publications were considered co-cited if they were both referenced by a third publication. In addition, bibliometrix was used to analyze publication output, citation performance, and the geographical distribution of research activity and to generate descriptive statistics and visual representations of bibliometric indicators across authors, institutions, and journals.

To evaluate author and journal productivity and citation impact, several complementary bibliometric indices were calculated. The h-index reflects the largest number h such that h publications have each received at least h citations, capturing both productivity and citation impact in a single measure.¹⁶ The g-index, proposed by Egghe¹⁷ as a citation-sensitive refinement of the h-index, is used less frequently in the literature than the h-index; it is defined as the largest number g such that the top g publications have collectively received at least g^2 citations. Because it incorporates the actual citation counts of the most-cited papers rather than simply counting papers above a fixed citation threshold, the g-index is always greater than or equal to the h-index and provides additional discriminatory power when comparing highly cited authors or journals.¹⁷ The m-index ($m = h/n$, where n is the number of years since the author's first publication) was also calculated to adjust productivity and impact for career duration, enabling more balanced comparisons among researchers at different career stages.¹⁶

Statistical Analysis

The bibliometric data were analyzed using descriptive statistical methods to summarize the literature's fundamental characteristics, including annual publication growth, document types, and

geographical distribution. Publication trends and citation performance were quantified using frequency and percentage distributions. Network visualizations, including co-authorship, co-citation, and keyword co-occurrence maps, were constructed using VOSviewer (version 1.6.20) and the Biblioshiny interface of the R-bibliometrix package to illustrate the field's collaborative and intellectual structure. Data cleaning and the generation of descriptive tables and charts were performed using Microsoft Excel (Microsoft Corp., Redmond, WA, USA). Since this study relied on the automated extraction of bibliographic metadata and focused on relational network mapping, no manual content coding or inferential statistical testing was required.

RESULTS

Overview and Annual Publication

The WoSCC retrieved 581 documents related to arthrocentesis (1989-2024), published across 54 sources, including journals, books, and conference proceedings. Original research articles constituted the majority of the dataset (n=485; 83.4%), followed by review articles (n=88; 15.1%). The overall annual growth rate of publications was 10.21%, with an average of 22.56 citations per article and a cumulative total of 7,505 citations. The dataset comprised contributions from 1,919 authors, reflecting a highly collaborative research environment. In total, 809 keywords plus terms and 746 author keywords were identified, covering core topics such as arthrocentesis, TMJ disorders, and arthroscopy.

The annual distribution of publications and citations has shown a steady upward trend since 2005, with annual output exceeding 20 papers from 2017 onward and peaking in 2024 (n=60). The highest mean citation rate was observed in 1991 (229.5 citations per article). The highest average citation rate since the 2000s was observed in 2015, with 49.16 citations per article. Overall, the data indicate that interest in arthrocentesis has remained consistently strong over the years, with steady increases in publication activity and research interest.

Analysis of Authors

The 25 most influential authors in TMJ arthrocentesis research are summarized in Table 1. Manfredini D. was the most prominent contributor, with an h-index of 15, g-index of 19, and m-index of 0.789, accumulating 751 citations across 19 publications since 2007. Guarda-Nardini L. ranked second, with an h-index of 13, a g-index of 17, and 647 citations from 17 publications. Dolwick M. F. had with an h-index of 12 and a g-index of 13, with 1,054 citations from 13 publications published since 1990.

The co-authorship network of authors with at least three publications is shown in Figure 2A. A total of 143 authors were distributed across 47 clusters, indicating a highly fragmented but centralized collaboration structure. The cluster led by Hamada exhibited the highest collaboration density (ten authors), followed by the Guarda-Nardini cluster (six authors). Co-citation analysis (Figure 2B), including authors with at least 25 citations, identified five clusters comprising 143 authors. Manfredini emerged as the most central and influential author, characterized by both high citation counts and high total link strength, while Guarda-Nardini showed similar co-citation prominence. Foundational authors, such as Nitzan, occupied intermediary positions, connecting multiple research themes within the arthrocentesis literature. Institutional

Table 1. Top twenty-five authors by productivity and impact in temporomandibular joint arthrocentesis research

Author	h-index	g-index	m-index	TC	NP	PY_start
Manfredini D	15	19	0.789	751	19	2007
Guarda-Nardini L	13	17	0.684	647	17	2007
Dolwick Mf	12	13	0.333	1054	13	1990
Hamada Y	11	15	0.478	240	15	2003
Kondoh T	10	11	0.385	221	11	2000
Seto K	10	11	0.385	227	11	2000
Emshoff R	9	10	0.346	263	10	2000
Ferronato G	9	9	0.5	418	9	2008
Holmlund Ab	9	12	0.321	256	12	1998
Segami N	8	10	0.258	413	10	1995
Dimitroulis G	7	9	0.226	413	9	1995
Grossmann E	7	12	0.538	145	12	2013
Kaneyama K	7	8	0.28	239	8	2001
Kurita K	7	7	0.241	182	7	1997
Nitzan Dw	7	7	0.194	904	7	1990
Gonzalez-Garcia R	6	8	0.3	237	8	2006
Nagori Sa	6	7	0.667	95	7	2017
Rudisch A	6	6	0.231	162	6	2000
Sato J	6	7	0.25	185	7	2002
Sembronio S	6	6	0.3	152	6	2006
Senturk Mf	6	8	0.6	109	8	2016
Iizuka T	5	5	0.161	293	5	1995
Jose A	5	5	0.625	90	5	2018
Kutuk N	5	5	0.625	42	5	2018
Machon V	5	5	0.333	134	5	2011

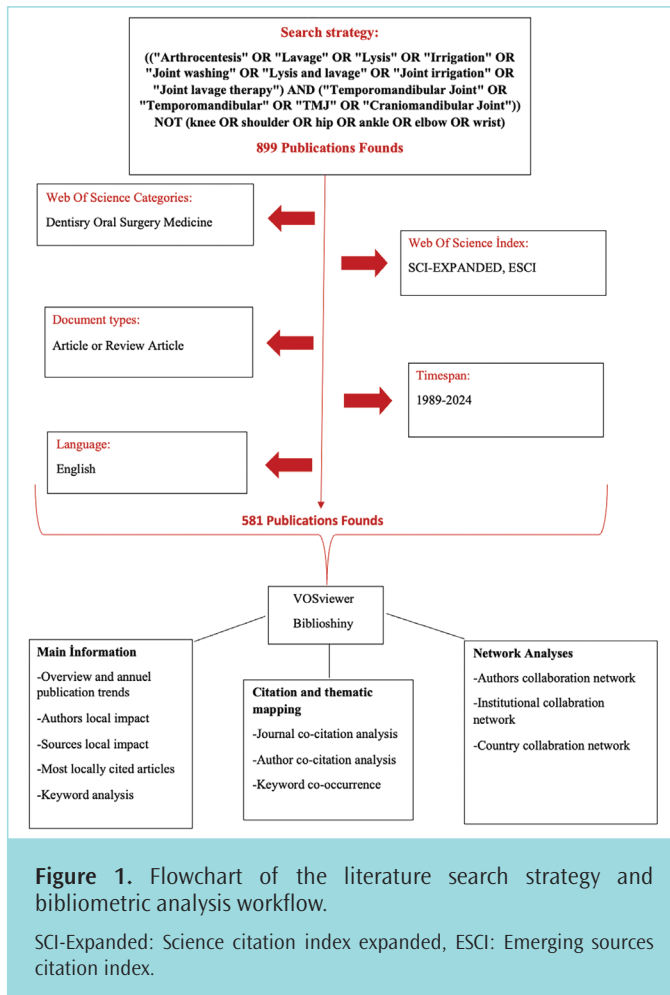
NP: Number of publications, TC: Total citations, PY_start: Publication year starting.

collaboration analysis (Figure 2C) identified Tsurumi University as the most productive institution with 18 publications, followed by Karolinska Institutet and the University of Padua with 16 and 14 publications, respectively. Network analysis revealed Karolinska Institutet as the most central collaborative institution, whereas the University of Padua, despite its high publication output, demonstrated relatively limited inter-institutional collaboration.

At the country level (Figure 2D), 30 countries were grouped into seven collaborative clusters. The United States exhibited the strongest international collaboration, particularly with Sweden, Japan, and Brazil, underscoring its central role in global TMJ arthrocentesis research. In contrast, Türkiye, despite its high publication output, showed lower total link strength, suggesting predominantly domestic collaboration patterns.

Analysis of Sources

Table 2 presents the ten most productive and influential journals in TMJ arthrocentesis research. The Journal of Oral and Maxillofacial Surgery ranked first with 118 publications and 4,061 citations (h-index =37), followed by the International Journal of Oral and Maxillofacial Surgery and the Journal of Cranio-Maxillofacial Surgery, with 66 and 54 publications, respectively.



Journal co-citation analysis (Figure 3A) identified 21 journals connected through 174 citation links, forming six distinct clusters. Node size corresponded to publication volume. The Journal of Oral and Maxillofacial Surgery occupied a central position within the network, reflecting both high publication output and strong citation connectivity. Notably, strong interconnections among the Journal of Oral and Maxillofacial Surgery, the International Journal of Oral and Maxillofacial Surgery, and the Journal of Cranio-Maxillofacial Surgery suggest that these journals function as key interdisciplinary bridges within the field.

Analysis of Most Local Cited Articles

Table 3 presents the local citation (LC) analysis of TMJ arthrocentesis-related publications and lists the ten articles with the highest LC counts. The top-ranked article by Manfredini D (2012, Journal of Oral Rehabilitation) received 48 LCs and 84 global citations, yielding an LC/GC ratio of 57.14%. This was followed by studies by Schiffman E (2017, Journal of Dental Research) and Guarda-Nardini L (2015, Journal of Oral Rehabilitation). Additional highly cited contributions included studies by Nagori SA (2018) and Kropmans TJB (1999). LC/GC ratios ranged from 1.69% to 71.43%, indicating substantial variation in the relative local influence of seminal publications within the arthrocentesis literature.

Keyword Analysis

Keyword co-occurrence analysis (Figure 3B) identified the most frequent terms: "arthrocentesis", "temporomandibular joint",

"temporomandibular joint disorders", "arthroscopy", "hyaluronic acid", "temporomandibular disorders", and "osteoarthritis". Larger nodes reflect higher frequency of occurrence, while shorter distances and thicker links indicate stronger associations between keywords.

Temporal keyword analysis (Figure 3C) revealed a clear shift in research focus. Earlier studies primarily emphasized mechanical and procedural concepts, whereas publications after 2020 increasingly focused on biological and regenerative approaches, particularly platelet-rich plasma (PRP) and hyaluronic acid. This trend indicates a transition from symptom-oriented management toward biologically driven strategies aimed at joint and disc repair. In contrast, the relatively stable or declining prominence of terms such as "occlusal splints" suggests their continued clinical relevance but reduced novelty in recent research.

Collaboration Network Analysis

Collaboration networks among authors, institutions, and countries are presented in Figure 4. Author-level analysis revealed a centralized collaboration structure dominated by several core clusters. The red cluster, led by Manfredini, Guarda-Nardini, and Ferronato, formed the most densely interconnected group, indicating sustained scholarly collaboration. The blue cluster was characterized by Hamada's central bridging role, linking multiple clusters through collaborations with authors such as Holmlund, Gynther, and Takahashi. The orange cluster, including Sato, Segami, and Kaneyama, also demonstrated strong internal collaboration among its members.

Analysis of institutional collaboration showed that the most intensive network consisted primarily of Japanese institutions (Tsurumi University, Nihon University, and Tohoku University) together with the Karolinska Institute. A secondary cluster included Atatürk University and Akdeniz University, reflecting regionally concentrated yet productive institutional partnerships.

At the country level, the United States occupied the most central position within the global collaboration network, maintaining strong collaborative ties with Türkiye, Japan, and China. European countries demonstrated moderate connectivity, whereas many Asian and Latin American countries exhibited more regionally confined patterns of collaboration. Overall, these findings illustrate a globally distributed yet regionally stratified structure of scientific collaboration in TMJ arthrocentesis research.

DISCUSSION

To the best of our knowledge, this study is the first comprehensive bibliometric analysis to focus exclusively on TMJ arthrocentesis, covering publications from 1989 to 2024. The primary aim was to quantitatively map the scientific development, thematic evolution, and intellectual structure of research in this field. Analysis of 581 publications revealed several key findings: a sustained increase in research output since 2005; a concentration of publications within a limited number of core journals, most notably the Journal of Oral and Maxillofacial Surgery; a dominance of influential authors such as Manfredini and Guarda-Nardini; and a clear thematic shift toward regenerative and biologically oriented treatment strategies.

The purpose of this bibliometric analysis is to highlight the fragmented structure of the existing literature, identify knowledge gaps, and clarify how the arthrocentesis technique has evolved over time.

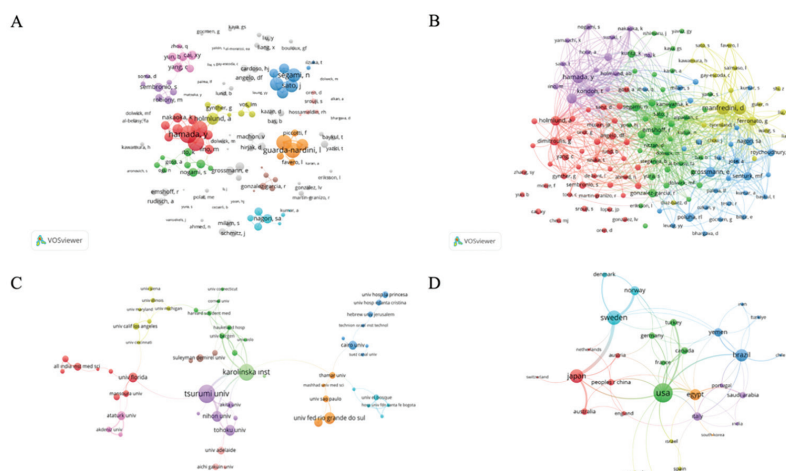


Figure 2. Collaboration and co-citation networks in TMJ arthrocentesis research visualized using VOSviewer. (A) Author co-authorship network showing collaborative relationships among authors with at least three publications. Node size represents publication output, while link thickness indicates collaboration strength. (B) Author co-citation network illustrating the intellectual structure of the field based on co-citation relationships. Larger nodes reflect higher citation impact, and clusters represent closely related research communities. (C) Institutional collaboration network depicting inter-institutional research partnerships. Node size corresponds to publication volume, and links indicate collaborative connections between institutions. (D) Country collaboration network demonstrating international research cooperation patterns in TMJ arthrocentesis. Node size reflects publication output, and link strength represents the intensity of collaboration between countries. Colors indicate distinct collaboration clusters.

TMJ: Temporomandibular joint.

Table 2. The ten most productive and influential journals in temporomandibular joint arthrocentesis research

Source	h-index	g-index	m-index	TC	NP	PY_start
Journal of Oral and Maxillofacial Surgery	37	52	0.86	4061	118	1990
International Journal of Oral and Maxillofacial Surgery	29	36	0.78	1990	66	1997
Journal of Cranio-Maxillofacial Surgery	23	34	0.83	1286	54	2000
Oral Surgery Oral Medicine Oral Pathology Oral Radiology	23	32	0.61	1580	34	1998
British Journal of Oral and Maxillofacial Surgery	13	29	0.54	984	34	1998
Journal of Oral Rehabilitation	14	22	0.58	888	26	1985
Cranio: The Journal of Craniomandibular & Sleep Practice	14	18	0.48	487	24	1994
Journal of Maxillofacial and Oral Surgery	10	18	0.83	574	21	2004
Journal of Oral & Facial Pain and Headache	8	10	0.4	272	17	2010
Oral and Maxillofacial Surgery Heidelberg	8	10	0.3	220	16	2010

NP: Number of publications, TC: Total citations, PY_start: Publication year starting.

Table 3. The ten most locally cited articles in temporomandibular joint arthrocentesis research

Document	DOI	Year	LC	GC	LC/GC ratio(%)	NLC	NGC
Manfredini D, 2012, J Oral Rehabil	10.1111/j.1365-2842.2011.02265.x	2012	48	84	57.14	21.00	3.21
Schiffman El, 2007, J Dent Res	10.1177/154405910708600109	2007	23	85	27.06	10.00	2.32
Guarda-Nardini L, 2015, J Oral Rehabil	10.1111/joor.12282	2015	21	42	50.00	9.73	0.85
Nagori Sa, 2018, J Oral Rehabil	10.1111/joor.12665	2018	17	29	58.62	17.41	1.52
Kropmans Tjb, 1999, J Oral Rehabil	10.1046/j.1365-2842.1999.00417.x	1999	16	37	43.24	9.00	1.30
Ozdamar Sm, 2017, J Oral Rehabil	10.1111/joor.12467	2017	15	23	65.22	21.18	0.97
Al-Baghdadi M, 2014, J Dent Res	10.1177/0022034514528333	2014	14	76	18.42	6.50	2.87
Haigler Mc, 2018, J Am Dent Assoc	10.1016/j.adaj.2018.07.025	2018	14	47	29.79	14.34	2.46
Toameh Mh, 2019, Dent Med Probl	10.17219/dmp/109329	2019	14	25	56.00	16.10	2.63
Guarda-Nardini L, 2011, J Oral Rehabil	10.1111/j.1365-2842.2010.02164.x	2011	12	23	52.17	13.50	1.08

LC: Local citations, GC: Global citations, LC/GC ratio: Percentage of local to global citations, NLC: Normalized local citations, NGC: Normalized global citations.

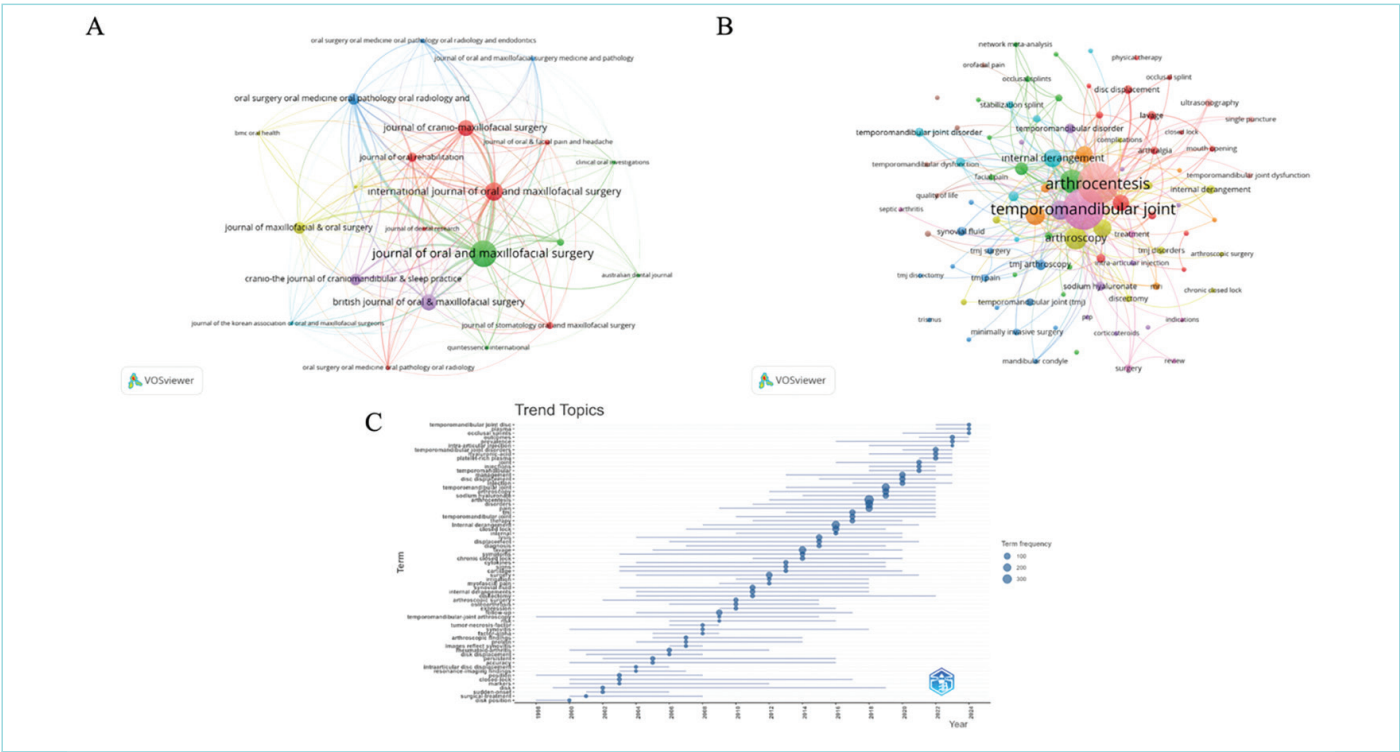


Figure 3. Journal, keyword, and thematic trend analyses in TMJ arthrocentesis research. (A) Journal co-citation network showing relationships among the most influential journals. (B) Keyword co-occurrence network illustrating the main research topics and their interrelationships. (C) Trend topics analysis demonstrating the temporal evolution of frequently used keywords over time.

TMJ: Temporomandibular joint.

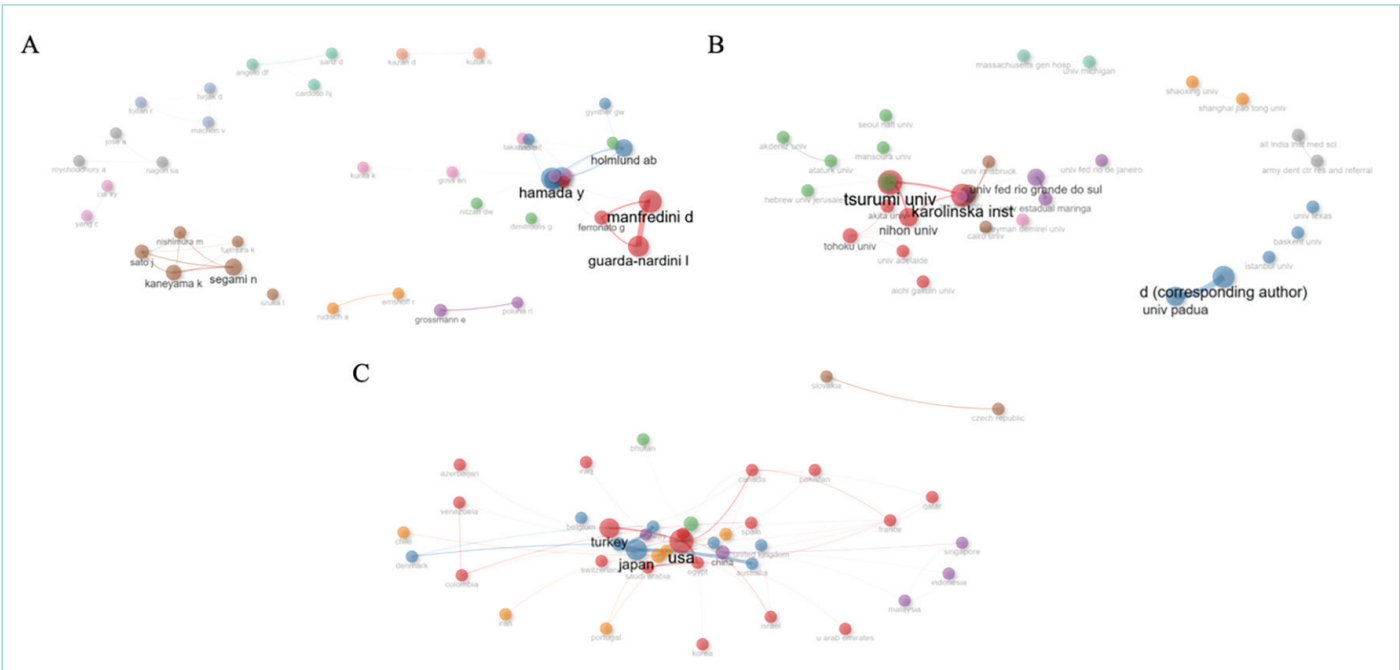


Figure 4. Global collaboration networks in TMJ arthrocentesis research (1989-2024). (A) Author Network: Visualizes co-authorship clusters, highlighting the central roles of Manfredini, Guarda-Nardini, and Hamada. (B) Institutional Network: Illustrates research partnerships among leading academic centers, with Tsurumi University and Karolinska Institutet appearing as major hubs. (C) Country Network: Displays international cooperation patterns, showing the United States, Türkiye, and Japan as the primary contributors to global scientific exchange.

TMJ: Temporomandibular joint.

The annual publication trends show increases and decreases that correspond to major developments in the field. After Nitzan first described arthrocentesis in 1991, studies of hyaluronic acid and steroid injections appeared in the early 2000s. Ultrasound-guided techniques emerged around 2010, and in 2012, arthrocentesis was defined as a minimally invasive, effective, and low-risk treatment for TMJ disorders. In the same year, Şentürk and Cambazoglu⁷ proposed a classification of arthrocentesis techniques based on the number of puncture sites, distinguishing single-puncture from double-puncture arthrocentesis and providing a standardized framework that has since been widely used to compare procedural variants. From 2015 onward, regenerative approaches such as PRP and stem-cell-based products gained attention; after 2020, efforts shifted toward standardization and improving levels of evidence. These milestones explain the overall publication pattern.¹⁸⁻²⁴

The development of research on TMJ arthrocentesis appears to mirror the scientific trends observed in studies of other joints. In the context of TMJ arthrocentesis, this procedure offers direct, albeit rare, access to synovial fluid, and its biochemical profiling has the potential to substantially advance personalized joint-specific treatment strategies. A comprehensive 2024 bibliometric analysis of synovial fluid research in osteoarthritis found a marked increase over the past decade in studies focusing on the diagnostic and biological significance of synovial fluid. This global trend aligns with the rise in TMJ arthrocentesis publications after 2018, highlighting growing interest in minimally invasive treatment approaches. While large joints such as the knee and hip have been extensively investigated—particularly regarding synovial fluid biomarkers, inflammatory pathways, and intra-articular therapies—research on the TMJ remains relatively limited. This indicates that further clinical and scientific work is needed and that the present study not only represents the first bibliometric analysis in this field but also addresses a meaningful gap in the current literature.²⁵⁻²⁷

The most clinically relevant finding of this study is the thematic evolution revealed by keyword co-occurrence and trend analyses. Over the past three decades, TMJ arthrocentesis research has shifted from symptom-oriented pain relief toward biologically driven strategies aimed at repairing TMJ disc pathology. The increasing volume of publications on regenerative approaches, together with a growing emphasis on treatment outcomes, reflects a broader transition toward evidence-based comparisons of different techniques. Recent high-level evidence supports this trend. Turosz et al.²⁸ compared arthrocentesis techniques and intra-articular biological agents exclusively through randomized controlled trial-based systematic reviews, identifying approaches supported by stronger evidence. In parallel, Zhang et al.²⁹ systematically compared regenerative agents such as PRP, i-PRF, and bone marrow aspirate concentrate (BMAC) across different follow-up periods, providing robust comparative data on their clinical effectiveness. Collectively, these studies reinforce the growing focus on regenerative treatments and align closely with the patterns identified in our bibliometric analysis. However, the absence of standardized application protocols, including preparation methods, injection volume, treatment frequency, and follow-up duration, continues to generate substantial heterogeneity in reported outcomes, highlighting a critical gap in the current literature.

Analysis of the literature's intellectual structure shows that research on TMJ arthrocentesis is primarily shaped by studies focused on oral and maxillofacial surgery published in high-impact journals, such

as the *Journal of Oral and Maxillofacial Surgery*. Core publications by Manfredini, Guarda-Nardini, and Nagori form the main scientific foundation of the field, as reflected by their high LC rates. In contrast, the work of Schiffman et al.³⁰ demonstrates broader influence beyond arthrocentesis-specific research, with high global citation counts driven by its relevance to orofacial pain and multidisciplinary TMD research. This pattern indicates that while arthrocentesis research is anchored in surgical literature, its conceptual development benefits from interdisciplinary contributions, particularly from pain-related disciplines.

The distinct separation of clusters observed in the visualizations indicates strong regional research hubs in the TMJ arthrocentesis literature; however, international collaboration among these hubs appears limited. Notably, Italian and Japanese clusters form the core of the network, whereas smaller clusters from other countries remain relatively isolated. This pattern suggests that the field continues to evolve around country-based research foci, and that strengthening multicenter and international collaborations may enhance the coherence, methodological diversity, and overall evidence base of the literature.

Analysis of the geographical distribution of publications demonstrates a pronounced global interest in TMJ arthrocentesis, with the United States, Türkiye, and Japan as the leading contributors. Türkiye's substantial research output reflects a sustained clinical commitment to minimally invasive TMJ interventions, facilitated by the procedural simplicity of arthrocentesis and by its low morbidity profile, rapid postoperative recovery, and limited equipment requirements. This productivity is further reinforced by the country's large young demographic, the rising prevalence of stress-related bruxism, and the longstanding integration of regenerative therapies into routine clinical practice.²² Although Türkiye is among the most productive countries in the TMJ arthrocentesis literature, its low total link strength may stem from research being conducted largely within the same institutions or domestic networks. This limited integration into international research structures likely weakens the country's overall capacity for collaboration. Japan's prominent position is consistent with its historical scientific leadership in the field, grounded in foundational contributions such as Murakami's double-needle arthrocentesis technique, Honda and Kurita's systematization of operative arthroscopy, and Yasunori Yamamoto's endoscopic characterization of intra-articular anatomy.³¹ The high burden of age-related degenerative joint conditions in Japan—driven by the world's oldest population—also likely stimulates greater demand for minimally invasive TMJ treatments.³² In the United States, the strong scientific engagement with arthrocentesis is attributable not only to the epidemiological prevalence of TMJ disorders but also to robust institutional research infrastructures and funding ecosystems. Notably, the National Institutes of Health routinely supports investigations into TMJ pathophysiology, minimally invasive surgical strategies, and regenerative modalities under thematic programs such as “pain disorders”, “musculoskeletal degeneration”, and “oral and maxillofacial health”, thereby sustaining a high level of scholarly activity in arthrocentesis and arthroscopy.³³

The primary strength of this study lies in its novelty as the first comprehensive bibliometric analysis dedicated to TMJ arthrocentesis research. By integrating performance analysis with Biblioshiny and scientific mapping with VOSviewer, the study provides a multi-level evaluation of research productivity, collaborative networks, and the

intellectual structure of the field, thereby extending beyond purely descriptive bibliometric metrics.

Study Limitations

This study has several limitations. First, data retrieval was restricted to the WoSCC. Although WoSCC is widely recognized as a high-quality citation database, the exclusion of other major sources, such as Scopus and Embase, may have led to the omission of relevant publications. Second, the time-dependent nature of citation accumulation inherently disadvantages recently published studies, because they have had limited opportunity to accrue citations, potentially underestimating their scientific impact. Third, the methodological framework relied exclusively on VOSviewer and Biblioshiny. Although both tools are effective for performance analysis and network visualization, an alternative tool, CiteSpace-particularly useful for identifying research-front “burst” topics- was not employed and might have provided additional insights. This study shares the inherent limitations of bibliometric analyses. Quantitative indicators, such as publication counts or total citations, do not necessarily reflect methodological quality or clinical relevance, and highly cited studies are not always synonymous with high scientific merit, as citation frequency may also be influenced by controversy or negative attention. Author-level productivity figures in Table 1 reflect the broad, multi-term search strategy adopted in this study (arthrocentesis, lavage, lysis, irrigation, and related terms). Consequently, author-specific publication counts may differ from those obtained using a narrower, single-term search in other databases such as PubMed.

CONCLUSION

This study provides the first comprehensive bibliometric overview of TMJ arthrocentesis research published between 1989 and 2024. Analysis of 581 publications reveals a clear thematic shift from symptom-oriented pain management toward biologically driven and regenerative treatment strategies, particularly involving PRP, BMAC, and hyaluronic acid. Research output is concentrated in a limited number of specialized maxillofacial surgery journals and reflects broader regenerative trends observed in other joints, with increasing emphasis on methodological rigor and evidence-based comparisons. The growing adoption of regenerative agents underscores the need for standardized application protocols and high-quality comparative studies to enhance clinical translation. Overall, this bibliometric analysis offers a concise reference for clinicians and researchers to understand the evolving research landscape of TMJ arthrocentesis and guide future investigations in this field.

MAIN POINTS

- This study is the first comprehensive bibliometric analysis of temporomandibular joint (TMJ) arthrocentesis from 1989 to 2024.
- Research activity in the field has shown a steady upward trend since 2005, culminating in a significant peak in publication output in 2024.
- A major thematic shift has been identified: a movement from traditional mechanical irrigation concepts toward regenerative therapeutic modalities, such as platelet-rich plasma and hyaluronic acid.

- The United States, Türkiye, and Japan are the leading contributors to the global scientific output on TMJ arthrocentesis.
- As the field grows, standardized application protocols are critically needed to reduce heterogeneity in clinical outcomes.

ETHICS

Ethics Committee Approval: This study was based exclusively on data retrieved from the Web of Science Core Collection database and did not involve human participants, animals, or patient records. Therefore, ethical approval and informed consent were not required.

Informed Consent: Not applicable, as this study is a bibliometric analysis of previously published literature.

Footnotes

Authorship Contributions

Concept: M.I., A.A., Design: M.I., A.A., Data Collection and/or Processing: A.A., Analysis and/or Interpretation: M.I., A.A., Literature Search: M.I., A.A., Writing: M.I., A.A.

DISCLOSURES

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

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

Declaration of Generative AI and AI-assisted Technologies in the Writing Process: During the preparation of this work, the author(s) used Claude (Anthropic) to improve readability and language. After using this tool/service, the author(s) reviewed and edited the content as needed and took full responsibility for the content of the publication.

REFERENCES

1. Kaneda M, Shimpo Y, Yoshida K, Kubo R, Kobayashi F, Mishima A, et al. Structural features of the temporomandibular joint evaluated by mri and their association with oral function and craniofacial morphology in female patients with malocclusion: a cross-sectional study. *J Clin Med*. 2025; 14(14): 4921.
2. Małgorzata P, Małgorzata KM, Karolina C, Gala A. Diagnostic of temporomandibular disorders and other facial pain conditions-narrative review and personal experience. *Medicina (Kaunas)*. 2020; 56(9): 472.
3. Zhang JM, Yun J, Zhou TQ, Zhang Y, Gao C. Arthrocentesis for temporomandibular joint disorders: a network meta-analysis of randomised controlled trials. *BMC Oral Health*. 2024; 24(1): 1108.
4. Sá M, Faria C, Pozza DH. Conservative versus invasive approaches in temporomandibular disc displacement: a systematic review of randomized controlled clinical trials. *Dent J (Basel)*. 2024; 12(8): 244.
5. Mauro G, Verdecchia A, Suárez-Fernández C, Nocini R, Mauro E, Zerman N. Temporomandibular disorders management-what's new? A scoping review. *Dent J (Basel)*. 2024; 12(6): 157.
6. Nitzan DW, Dolwick MF, Martinez GA. Temporomandibular joint arthrocentesis: a simplified treatment for severe, limited mouth opening. *J Oral Maxillofac Surg*. 1991; 49(11): 1163-7; discussion 1168-70.
7. Şentürk MF, Cambazoğlu M. A new classification for temporomandibular joint arthrocentesis techniques. *Int J Oral Maxillofac Surg*. 2015; 44(3): 417-8.
8. Siewert-Gutowska M, Pokrowiecki R, Kamiński A, Zawadzki P, Stopa Z. State of the art in temporomandibular joint arthrocentesis-a systematic review. *J Clin Med*. 2023; 12(13): 4439.

9. Ninkov A, Frank JR, Maggio LA. Bibliometrics: methods for studying academic publishing. *Perspect Med Educ.* 2022; 11(3): 173-6.
10. Cascajares M, Alcayde A, Salmerón-Manzano E, Manzano-Agugliaro F. The bibliometric literature on scopus and wos: the medicine and environmental sciences categories as case of study. *Int J Environ Res Public Health.* 2021; 18(11): 5851.
11. Al-Moraissi EA, Christidis N, Ho YS. Publication performance and trends in temporomandibular disorders research: a bibliometric analysis. *J Stomatol Oral Maxillofac Surg.* 2023; 124(1): 101273.
12. Xu C, Yu C, Wang T, Gao X. Bibliometric and visualization analysis of temporomandibular joint osteoarthritis from 2004 to 2024. *Front Oral Health.* 2025; 6: 1596551.
13. Ferrillo M, Gallo V, Lippi L, Bruni A, Montrella R, Curci C, et al. The 50 most-cited articles on temporomandibular disorders: a bibliometric analysis. *J Back Musculoskelet Rehabil.* 2023; 36(2): 279-97.
14. Al-Sharaee Y, Al-Moraissi EA, Christidis N, Galvão EL, Falci SGM. Top 100 cited publications in the field of temporomandibular disorders: a bibliometric analysis. *Front Oral Health.* 2022; 3: 864519.
15. Wang Q, Jia J, Zhou C, Ye W, Bi R. A bibliometric analysis of research on temporomandibular joint disc displacement from 1992 to 2022. *Healthcare (Basel).* 2023; 11(14): 2108.
16. Hirsch JE. An index to quantify an individual's scientific research output. *Proc Natl Acad Sci U S A.* 2005; 102(46): 16569-72.
17. Egghe L. Theory and practise of the g-index. *Scientometrics.* 2006; 69(1): 131-52.
18. Nitzan DW, Dolwick MF, Heft MW. Arthroscopic lavage and lysis of the temporomandibular joint: a change in perspective. *J Oral Maxillofac Surg.* 1990; 48(8): 798-801; discussion 802.
19. Hepguler S, Akkoc YS, Pehlivan M, Ozturk C, Celebi G, Saracoglu A, et al. The efficacy of intra-articular sodium hyaluronate in patients with reducing displaced disc of the temporomandibular joint. *J Oral Rehabil.* 2002; 29(1): 80-6.
20. Monje-Gil F, Nitzan D, González-García R. Temporomandibular joint arthrocentesis. Review of the literature. *Med Oral Patol Oral Cir Bucal.* 2012; 17(4): e575-81.
21. Dayisoğlu EH, Cifci E, Uckan S. Ultrasound-guided arthrocentesis of the temporomandibular joint. *Br J Oral Maxillofac Surg.* 2013; 51(7): 667-8.
22. Hancı M, Karamese M, Tosun Z, Aktan TM, Duman S, Savacı N. Intra-articular platelet-rich plasma injection for the treatment of temporomandibular disorders and a comparison with arthrocentesis. *J Craniomaxillofac Surg.* 2015; 43(1): 162-6.
23. Cömert Kiliç S, Güngörmüş M. Is arthrocentesis plus platelet-rich plasma superior to arthrocentesis plus hyaluronic acid for the treatment of temporomandibular joint osteoarthritis: a randomized clinical trial. *Int J Oral Maxillofac Surg.* 2016; 45(12): 1538-44.
24. Simplot TM, Fidele NB, Augustin MM, Grace PL, Pierre MM, Julia ZM, et al. Controversies in the management of arthrocentesis treatment of temporomandibular joint disorders: systematic review. *J Med Life.* 2025; 18(8): 721-31.
25. Han G, Cai L, Li B, Li Q, Yue Luo, Wang Q, et al. Bibliometric analysis of synovial in osteoarthritis in the last 10 years. *Heliyon.* 2024; 10(13): e33406.
26. Haraden CA, Huebner JL, Hsueh MF, Li YJ, Kraus VB. Synovial fluid biomarkers associated with osteoarthritis severity reflect macrophage and neutrophil related inflammation. *Arthritis Res Ther.* 2019; 21(1): 146.
27. Monibi F, Roller BL, Stoker A, Garner B, Bal S, Cook JL. Identification of synovial fluid biomarkers for knee osteoarthritis and correlation with radiographic assessment. *J Knee Surg.* 2016; 29(3): 242-7.
28. Turosz N, Chęcińska K, Chęciński M, Lubecka K, Bliźniak F, Chlubek D, et al. Temporomandibular joint injections and lavage: an overview of reviews. *J Clin Med.* 2024; 13(10): 2855.
29. Zhang JM, Yun J, Zhou TQ, Zhang Y, Gao C. Arthrocentesis for temporomandibular joint disorders: a network meta-analysis of randomised controlled trials. *BMC Oral Health.* 2024; 24(1): 1108.
30. Schiffman EL, Look JO, Hodges JS, Swift JQ, Decker KL, Hathaway KM, et al. Randomized effectiveness study of four therapeutic strategies for TMJ closed lock. *J Dent Res.* 2007; 86(1): 58-63.
31. Murakami K. Rationale of arthroscopic surgery of the temporomandibular joint. *J Oral Biol Craniofac Res.* 2013; 3(3): 126-34.
32. Wu Z, Lin Y, Lin Y. Impact of age on degenerative joint disease of the temporomandibular joint: a systematic review and meta-analysis. *Medicine (Baltimore).* 2025; 104(17): e41915.
33. Temporomandibular Disorders | National Institute of Dental and Craniofacial Research [Internet]. Available from: <https://www.nidcr.nih.gov/grants-funding/funded-research/research-investments-advances/temporomandibular-disorders>

Analysis of the Relationship Between Disease Self-Management and Psychosocial Adjustment of Celiac Patients

✉ İlknur Bodur¹, ✉ Hilal Tüzer²

¹Clinic of Nursing, University of Health Sciences Türkiye, Dışkapı Yıldırım Beyazıt Training and Research Hospital, Ankara, Türkiye

²Department of Nursing, Ankara Yıldırım Beyazıt University Faculty of Health Sciences, Ankara, Türkiye

Abstract

BACKGROUND/AIMS: In this study, it was aimed to analyze the relationship between disease self-management and psychosocial adjustment of celiac patients.

MATERIALS AND METHODS: This study was designed as descriptive and correlational. The data for the study were collected from 279 individuals with celiac disease between 25 July 2024 and 25 October 2024. In our study, the psychosocial adjustment levels of patients with celiac disease were evaluated with respect to various socio-demographic factors (such as gender, marital status, income level, and education level).

RESULTS: Analyses indicated that female patients had worse psychosocial adjustment than male patients. Single patients had higher scores on the sub-dimensions of domestic environment and psychological distress than married patients. Participants who had another family member diagnosed with celiac disease demonstrated better healthcare orientation. Increasing age was associated with greater acceptance of chronic disease and better treatment compliance. The results reveal that factors such as gender, marital status, and income level are significant determinants of self-management and psychosocial adjustment among patients with celiac disease.

CONCLUSION: To improve the psychological adaptation and self-management among patients with celiac disease, support programs should consider factors such as income, gender, marital status, and age. Financial support should be provided to cope with the economic burden of the diet, and educational programs should be organized not only for patients but also for their families so that they can be supportive in this process.

Keywords: Celiac disease, chronic disease, self-management, psychosocial adjustment

INTRODUCTION

Celiac disease is a chronic, immune-mediated disorder that develops in genetically predisposed individuals following exposure to gluten. Gluten, a group of storage proteins containing prolamins found in wheat, rye, barley, and oats, triggers an abnormal immune response that leads to characteristic serological and histopathological changes in the small intestine.^{1,2} Celiac disease affects approximately 1% of the global population and may present at any age. Notably, its incidence

has increased substantially over the past four decades, likely due to improved diagnostic awareness and environmental factors.³

Gluten exposure in individuals with celiac disease results in immune-mediated damage to the small intestinal mucosa, causing villous atrophy and a reduction in absorptive surface area. Consequently, malabsorption of essential micronutrients such as iron, folate, vitamin B12, and fat-soluble vitamins frequently occurs. Diagnosis is typically

To cite this article: Bodur İ, Tüzer H. Analysis of the relationship between disease self-management and psychosocial adjustment of celiac patients. Cyprus J Med Sci. 2026;11(5):393-400

ORCID IDs of the authors: İ.B. 0000-0001-5511-1611; H.T. 0000-0002-9929-3688.



Corresponding author: İlknur Bodur

E-mail: ilknurbodur89@gmail.com

ORCID ID: orcid.org/0000-0001-5511-1611

Received: 16.04.2026

Accepted: 03.08.2026

Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.

This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

based on serological markers, including anti-gliadin antibodies, anti-tissue transglutaminase, and anti-endomysial antibodies, in conjunction with clinical findings. Histological confirmation of villous atrophy via small intestinal biopsy remains the diagnostic gold standard.⁴ Clinical manifestations vary widely and may include chronic diarrhea, abdominal distension, recurrent abdominal pain, constipation, anemia, osteoporosis, growth retardation, micronutrient deficiencies, and recurrent aphthous ulcers.²

Currently, the only effective treatment is strict adherence to a lifelong gluten-free diet, which alleviates symptoms, normalizes serological markers, and allows mucosal healing.¹ However, compliance with the diet requires continuous education and support. Dietitian-guided counseling helps patients learn how to interpret food labels, identify hidden sources of gluten, and prevent cross-contamination. Patient and family education is also critical to ensure safety in food preparation at home.²

The United States Centers for Disease Control and Prevention defines chronic disease as: “Chronic diseases are defined broadly as conditions that last 1 year or more and require ongoing medical attention or limit activities of daily living or both”.⁵

Self-management of chronic diseases refers to patients’ ability to cope with symptoms, adhere to treatment regimens, adopt lifestyle modifications, and address both physical and psychosocial consequences of their condition.⁶ Effective self-management not only improves health-related quality of life but also reduces the economic burden on healthcare systems.⁷ A prerequisite for self-management is patient empowerment, enabling individuals to actively participate in managing their condition.⁸

Psychosocial adjustment is defined as the ability to maintain a positive sense of self despite disease-related challenges. It is a dynamic process influenced by both individual and environmental factors.⁹ Successful adjustment implies that patients can adapt cognitively and emotionally to physical and functional changes. Conversely, maladjustment leads to adverse consequences such as depression, anxiety, psychosomatic symptoms, fatigue, social withdrawal, and impaired cognitive functioning.¹⁰

Given that celiac disease requires permanent lifestyle modifications, both self-management and psychosocial adjustment are of paramount importance. This study aimed to examine the relationship between chronic illness self-management and psychosocial adjustment among adults diagnosed with celiac disease.

Research Questions

Is there a significant relationship between chronic disease self-management skills and psychosocial adjustment in adults with celiac disease?

Do psychosocial adjustment and self-management levels differ according to socio-demographic and clinical characteristics (gender, marital status, income level, adherence to an additional diet, and presence of another family member with celiac disease)?

MATERIALS AND METHODS

Study Design

This study was a descriptive, correlational investigation examining the relationship between chronic disease self-management and psychosocial adjustment in adults diagnosed with celiac disease.

Population and Sample

The study population consisted of individuals registered with a celiac association in Türkiye. A priori sample size calculation indicated that a minimum of 234 participants was required to achieve a 95% confidence level. A total of 279 individuals who met the inclusion criteria and agreed to participate were included in the final analysis.

Inclusion and Exclusion Criteria

Inclusion criteria were as follows:

Having a confirmed diagnosis of celiac disease,

Being 18 years of age or older,

Volunteering to participate in the study,

Being able to communicate verbally.

Exclusion criteria included:

- Being under 18 years of age and
- Declining to participate in the study.

Data Collection Instruments

Data were collected using three instruments: a patient information form, the chronic illness self-management scale (CISM), and the psychosocial adjustment to illness scale-self-report (PAIS-SR).

The patient information form was developed by the researchers and included questions on participants’ socio-demographic and clinical characteristics.

Chronic illness self-management scale: The CISM was developed by Ngai et al.⁶ in 2020, and its Turkish validity and reliability were established by Öztürk et al.¹¹ in 2021. The scale consists of 21 items across four sub-dimensions: self-stigma, coping with stigma, healthcare efficiency, and compliance with treatment. Items are rated on a 5-point Likert scale. Scores are calculated using the arithmetic mean method, with higher scores indicating better self-management abilities.¹¹

The psychosocial adjustment to illness scale-self-report: Was originally developed by Derogatis and Lopez in 1983 to assess psychosocial adjustment to illness, and its Turkish validity and reliability were assessed by Adaylar in 1995.¹² The scale comprises 46 items grouped into seven sub-dimensions: healthcare orientation, vocational environment, domestic environment, sexual relationships, extended family relationships, social environment, and psychological distress. Each item is rated on a 4-point scale (0-3), with higher scores indicating poorer psychosocial adjustment. Total scores below 35 indicate good adjustment; scores between 35 and 51 indicate fair adjustment; scores above 51 indicate poor adjustment.¹²

Data Collection Procedure

Data were collected between July 25 and October 25, 2024, through face-to-face interviews and online questionnaires administered via Google Forms. Completion of the questionnaires required approximately 8-10 minutes. Participation was voluntary, and informed consent was obtained from all participants prior to data collection.

Ethical Considerations

Ethical approval for the study was obtained from the Ankara Yıldırım Beyazıt University Medical Research Ethics Committee (approval number: 06/810, date: 01.07.2024). The relevant association also granted institutional permission. Written informed consent was obtained from all participants, and permission to use the scales was secured from the authors.

Statistical Analysis

The research data were analyzed using IBM SPSS Statistics, version 26.0 (IBM Corp., Armonk, NY, USA). The descriptive findings in the study were reported as numbers, percentages, minimum and maximum values, means, standard deviations, and medians. The reliability of the scales was assessed according to Cronbach's alpha coefficient.¹³ The normality of distribution of the variables was examined according to whether the values of skewness and kurtosis were in the range of ± 1.5 , or not.¹⁴ For comparisons between two independent groups, the t-test or Mann-Whitney U test was used, and for variables with three or more groups, analysis of variance (ANOVA) or the Kruskal-Wallis test was used. When significant differences were detected by ANOVA or the Kruskal-Wallis test, the groups that differed were identified using Bonferroni-corrected multiple comparison tests. The relationship between continuous variables was measured using Pearson and Spearman correlation tests. A value of $p < 0.05$ was considered statistically significant in the analyses.

RESULTS

A total of 279 adults with celiac disease were included in the analysis. The majority of participants were female (88.5%), and the mean age was 33.29 ± 10.14 years. More than half of the participants (52.3%) were married. Nearly half (44.8%) held a bachelor's degree, and 45.9% reported that their income equaled their expenses. Most participants (92.5%) did not follow any diet in addition to a gluten-free diet, and

22.9% reported having at least one other family member diagnosed with celiac disease (Table 1).

The statistical and reliability findings for the scales used in the research are presented in Table 2. Of the participants, 118 (42.3%) had good psychosocial adjustment (< 35), 80 (28.7%) had fair psychosocial adjustment (35-51), and 81 (29.0%) had poor psychosocial adjustment (> 51).

As shown in Table 3, female participants had significantly higher Self-Stigma scores than male participants ($p = 0.015$). Marital status was significantly associated with scores on the self-stigma, coping with stigma, and compliance with treatment scales. Income level was significantly associated with scores on coping with stigma and healthcare efficiency. Participants following an additional diet had significantly higher self-stigma scores. Having another family member

Table 1. Descriptive findings of the participants

Characteristic	Category	n	%
Gender	Female	247	88.5
	Male	32	11.5
Age, years	Mean $\pm$ SD	33.29 $\pm$ 10.14	
Marital status	Married	146	52.3
	Single	133	47.7
Education level	Primary school	11	3.9
	High school	57	20.4
	Associate's degree	42	15.1
	Bachelor's degree	125	44.8
	Master's degree	44	15.8
Income level	Income less than expense	99	35.5
	Income equal to expense	128	45.9
	Income more than expense	52	18.6
Additional diet	No	258	92.5
	Yes	21	7.5
Family member with celiac disease	No	215	77.1
	Yes	64	22.9

SD: Standard deviation.

Table 2. Statistical and reliability findings of the scales

Scale/subscale	N	Min-max	Mean $\pm$ SD	Median	Cronbach's α
CISM self-stigma	279	1.00-4.71	2.35 $\pm$ 0.88	2.14	0.871
CISM coping with stigma	279	1.60-5.00	3.27 $\pm$ 0.72	3.20	0.760
CISM healthcare efficiency	279	1.75-5.00	4.16 $\pm$ 0.65	4.25	0.721
CISM compliance with treatment	279	1.00-5.00	4.39 $\pm$ 0.64	4.60	0.807
PAIS-SR healthcare orientation	279	0.00-22.00	10.21 $\pm$ 4.33	11.00	0.694
PAIS-SR vocational environment	279	0.00-18.00	4.66 $\pm$ 3.16	4.00	0.696
PAIS-SR domestic environment	279	0.00-20.00	6.40 $\pm$ 4.41	6.00	0.804
PAIS-SR sexual relationships	248	0.00-17.00	2.58 $\pm$ 3.38	1.00	0.875
PAIS-SR extended family relationships	279	0.00-12.00	3.88 $\pm$ 3.12	4.00	0.674
PAIS-SR social environment	279	0.00-17.00	5.27 $\pm$ 4.85	5.00	0.898
PAIS-SR psychological distress	279	0.00-21.00	8.37 $\pm$ 5.06	8.00	0.896
PAIS-SR total	279	4.00-111.00	41.11 $\pm$ 21.14	39.00	0.942

CISM: Chronic illness self-management scale, PAIS-SR: Psychosocial adjustment to illness scale-self report, SD: Standard deviation, Min-max: Minimum-maximum.

with celiac disease was associated with lower self-stigma scores and higher healthcare efficiency scores. Adherence to a gluten-free diet was significantly associated with scores for healthcare efficiency and compliance with treatment.

As shown in Table 4, PAIS-SR scores differed significantly according to several participant characteristics. Female participants scored higher than male participants on the healthcare orientation, domestic environment, social environment, and psychological distress subdimensions and on the total PAIS-SR score, indicating poorer psychosocial adjustment. Single participants had higher domestic environment and psychological distress scores and higher total PAIS-SR scores than married participants. Income level was significantly

associated with most PAIS-SR domains and the total score, with participants whose income was less than their expenses showing poorer adjustment. Participants following an additional diet had significantly higher scores on all PAIS-SR sub-dimensions and on the total score. In addition, participants without another family member with celiac disease had higher healthcare orientation scores; however, total PAIS-SR scores did not differ significantly according to family history.

As presented in Table 5, age was negatively correlated with self-stigma and positively correlated with coping with stigma, healthcare efficiency, and compliance with treatment, suggesting a tendency toward lower stigma and better self-management with increasing age. Age was also negatively associated with the PAIS-SR domestic environment

Table 3. Comparison of CISM subscale scores according to participant characteristics

Participant characteristics	SS mean $\pm$ SD	CS mean $\pm$ SD	HCE mean $\pm$ SD	CT mean $\pm$ SD
Gender				
Female	2.40 $\pm$ 0.88	3.25 $\pm$ 0.70	4.17 $\pm$ 0.65	4.38 $\pm$ 0.65
Male	2.00 $\pm$ 0.87	3.48 $\pm$ 0.84	4.15 $\pm$ 0.63	4.49 $\pm$ 0.58
Test statistic	2.446	-1.755	0.136	-0.968
p-value	0.015	0.080	0.892	0.333
Marital status				
Married	2.20 $\pm$ 0.86	3.37 $\pm$ 0.77	4.22 $\pm$ 0.64	4.49 $\pm$ 0.56
Single	2.53 $\pm$ 0.88	3.16 $\pm$ 0.64	4.10 $\pm$ 0.65	4.30 $\pm$ 0.72
Test statistic	-3.108	2.510	-1.319	-2.103
p-value	0.002	0.013	0.187	0.035
Income level				
Income less than expense ¹	2.46 $\pm$ 0.94	3.13 $\pm$ 0.73	3.96 $\pm$ 0.71	4.27 $\pm$ 0.71
Income equal to expense ²	2.33 $\pm$ 0.89	3.40 $\pm$ 0.70	4.26 $\pm$ 0.61	4.42 $\pm$ 0.65
Income more than expense ³	2.22 $\pm$ 0.73	3.25 $\pm$ 0.68	4.31 $\pm$ 0.51	4.58 $\pm$ 0.44
Test statistic	1.345	4.206	13.250	5.343
p-value	0.262	0.016	<0.001	0.069
Difference	-	1<2	1<2, 1<3	-
Additional diet				
No	2.29 $\pm$ 0.85	3.28 $\pm$ 0.73	4.17 $\pm$ 0.64	4.38 $\pm$ 0.65
Yes	3.16 $\pm$ 0.89	3.17 $\pm$ 0.60	4.07 $\pm$ 0.68	4.58 $\pm$ 0.48
Test statistic	-4.469	0.676	0.675	-1.202
p-value	<0.001	0.499	0.500	0.230
Family member with celiac disease				
No	2.43 $\pm$ 0.89	3.23 $\pm$ 0.69	4.12 $\pm$ 0.63	4.39 $\pm$ 0.65
Yes	2.11 $\pm$ 0.81	3.41 $\pm$ 0.79	4.32 $\pm$ 0.67	4.43 $\pm$ 0.63
Test statistic	2.571	-1.741	-2.339	-0.611
p-value	0.011	0.083	0.019	0.541
Adherence to a gluten-free diet				
Yes ¹	2.34 $\pm$ 0.87	3.30 $\pm$ 0.69	4.23 $\pm$ 0.58	4.45 $\pm$ 0.57
No ²	1.36 $\pm$ 0.10	2.90 $\pm$ 1.84	2.88 $\pm$ 1.24	3.80 $\pm$ 1.70
Sometimes ³	2.71 $\pm$ 1.01	2.91 $\pm$ 0.83	3.36 $\pm$ 0.80	3.67 $\pm$ 0.97
Test statistic	5.988	5.315	25.016	13.927
p-value	0.050	0.070	<0.001	0.001
Difference	-	-	3<1	3<1

Data are presented as mean $\pm$ standard deviation. Significant p-values are shown in bold. Superscript numbers indicate group codes used for post-hoc comparisons. For income level: 1= income less than expense, 2= income equal to expense, 3= income more than expense. For adherence to a gluten-free diet: 1= yes, 2= no, 3= sometimes.

CISM: Chronic illness self-management scale, SS: Self-stigma, CS: Coping with stigma, HCE: Healthcare efficiency, CT: Compliance with treatment, SD: Standard deviation.

Table 4. Comparison of PAIS-SR scores according to participant characteristics

Participant characteristic	Significant PAIS-SR domains	Direction of difference	PAIS-SR total mean $\pm$ SD	p-value
Gender	HCO, DE, SE, PD, total	Female participants had higher scores than males	Female: 42.43 $\pm$ 20.63; Male: 31.00 $\pm$ 22.58	0.004
Marital status	DE, PD, total	Single participants had higher scores than married participants	Married: 38.62 $\pm$ 21.46; Single: 43.86 $\pm$ 20.51	0.038
Income level	HCO, VE, DE, EFR, SE, PD, total	Participants whose income was less than expenses had poorer adjustment	Less: 48.89 $\pm$ 21.65; Equal: 38.48 $\pm$ 20.34; More: 32.81 $\pm$ 17.34	<0.001
Additional diet	HCO, VE, DE, SR, EFR, SE, PD, total	Participants following an additional diet had higher scores	No: 39.38 $\pm$ 19.79; Yes: 62.48 $\pm$ 25.69	<0.001
Family member with celiac disease	HCO	Participants without another affected family member had higher HCO scores	No: 41.96 $\pm$ 20.48; Yes: 38.28 $\pm$ 23.16	0.222

Only statistically significant PAIS-SR domains are presented in this table. Significant p-values are shown in bold. Higher scores indicate poorer psychosocial adjustment.

HCO: Healthcare orientation, VE: Vocational environment, DE: Domestic environment, SR: Sexual relationships, EFR: Extended family relationships, SE: Social environment, PD: Psychological distress, PAIS-SR: Psychosocial adjustment to illness scale-self report, SD: Standard deviation.

Table 5. Findings of correlation analysis between age, number of people living at home and duration of celiac disease and the scales

Scale/subscale	Age r (p)	Number of people living at home ¹ r (p)	Number of years lived with celiac disease ¹ r (p)
CISM self-stigma	-0.204 (0.001)	0.049 (0.414)	-0.087 (0.147)
CISM coping with stigma	0.139 (0.020)	-0.073 (0.221)	0.017 (0.781)
CISM healthcare efficiency	0.125 (0.037)	-0.096 (0.109)	0.195 (0.001)
CISM compliance with treatment	0.120 (0.045)	-0.040 (0.501)	0.020 (0.742)
PAIS-SR healthcare orientation	-0.056 (0.348)	0.056 (0.351)	-0.198 (0.001)
PAIS-SR domestic environment	-0.150 (0.012)	0.070 (0.241)	-0.107 (0.074)
PAIS-SR sexual relationships	0.187 (0.003)	0.015 (0.817)	-0.028 (0.659)
PAIS-SR extended family relationships	-0.112 (0.061)	0.142 (0.018)	0.007 (0.909)
PAIS-SR psychological distress	-0.270 (<0.001)	0.009 (0.884)	-0.118 (0.049)
PAIS-SR total	-0.110 (0.066)	0.070 (0.243)	-0.128 (0.033)

¹The Spearman correlation test. Significant p-values are shown in bold.

PAIS-SR: Psychosocial adjustment to illness scale-self report, CISM: Chronic illness self-management scale.

and psychological distress subscales, but positively associated with the sexual relationships subscale. The number of people living at home was positively correlated with only the PAIS-SR extended family relationships subscale. Longer disease duration was positively correlated with healthcare efficiency and negatively correlated with healthcare orientation, psychological distress, and total PAIS-SR scores, suggesting that participants with longer celiac disease duration reported better healthcare-related self-management and overall psychosocial adjustment.

Results of Correlation Analysis Between the Scales

Significant negative correlations were found between the self-stigma sub-dimension of the CISM and the coping with stigma, healthcare efficiency, and compliance with treatment sub-dimensions ($p < 0.05$). In contrast, coping with stigma, healthcare efficiency, and compliance with treatment were positively correlated with one another ($p < 0.05$). Significant positive correlations were also observed between the PAIS-SR subscales and the PAIS-SR total score ($p < 0.05$). In addition, the PAIS-SR total score was positively correlated with the CISM self-stigma sub-dimension ($r = 0.683$; $p < 0.001$) and negatively correlated with coping with stigma ($r = -0.487$; $p < 0.001$), healthcare efficiency ($r = -0.462$; $p < 0.001$), and compliance with treatment ($r = -0.241$; $p < 0.001$).

DISCUSSION

This study examined the relationship between chronic disease self-management and psychosocial adjustment in adults with celiac disease. The findings indicate that self-management capacity is closely associated with psychosocial outcomes. The mean PAIS-SR score suggested a moderate level of psychosocial adjustment, while nearly one-third of the participants demonstrated poor adjustment. This finding indicates that celiac disease should not be considered solely a gastrointestinal condition but also a chronic illness with substantial psychological, social, and behavioral consequences. This interpretation is supported by Rose et al.,¹⁵ who reported that adults living with celiac disease experience significant psychosocial challenges, including a changed identity, social restrictions, and the burden of maintaining a strict gluten-free diet.¹⁶ Similar psychosocial challenges have been reported in patients receiving chemotherapy, individuals with Behçet's disease, young adults with type 1 diabetes, and patients with heart failure.¹⁵⁻¹⁹ These similarities suggest that continuous disease management, uncertainty about disease progression, treatment-related restrictions, and lifestyle regulation may be common mechanisms affecting psychosocial adjustment across chronic illnesses.

Female patients scored higher than male patients on the healthcare orientation, domestic environment, social environment, and psychological distress subdimensions, as well as on the total PAIS-SR score, indicating poorer psychosocial adjustment. The poorer psychosocial adjustment observed among female participants may be related to gender-related social and domestic responsibilities. Women may experience a greater burden in managing dietary restrictions because they are often more involved in food preparation, caregiving, and family nutrition. However, given the predominance of female participants in the sample, this finding should be interpreted cautiously and not be generalized without further studies with a more balanced gender representation. Consistent with our findings, Türksen Ülkü and Ünsar²⁰ also reported that female patients with heart failure scored higher than males in the domestic environment, social environment, and psychological distress sub-dimensions and in the PAIS-SR total score.

Marital status also appeared to play an important role in psychosocial adjustment. Single participants showed poorer adjustment than married participants, particularly in the domestic environment and psychological distress, and in total PAIS-SR scores. Marriage may provide emotional support, shared responsibility, and practical assistance in maintaining dietary restrictions. Spousal support may also reduce the sense of isolation that can accompany chronic illness management. In contrast, single individuals may have fewer sources of daily support and need to manage the practical and emotional demands of celiac disease independently. However, in the study conducted by Türksen Ülkü and Ünsar²⁰ on patients with heart failure, single patients obtained higher scores than married ones from the domestic environment and social environment sub-scales, and married patients had higher scores than single ones in the sexual relationships sub-scale. In the study conducted by Senmar et al.²¹ on cancer patients, no statistically significant differences were found between single and married patients.

Income level was one of the most important factors associated with psychosocial adjustment. Participants whose expenses exceeded their income reported poorer adjustment across several domains. This may be explained by the economic burden of gluten-free products, which are often more expensive and less accessible than gluten-containing alternatives. Financial strain may limit access to safe foods, increase anxiety about dietary adherence, and restrict participation in social activities involving food. Therefore, the relationship between low income and poorer adjustment may reflect not only general socio-economic disadvantage but also the specific financial demands of celiac disease management. This finding highlights the need for economic and social policy interventions, such as financial support for gluten-free products and improved access to safe foods. In their study on elderly patients with heart failure, Vural Doğru and Karadakovan²² concluded that patients with poor income status obtained higher scores than patients with good income status from the healthcare orientation, domestic environment and extended family relationships sub-scales. In the study conducted by Çam and Nehir²³ on patients who experienced myocardial infarction, it was detected that the PAIS-SR total score obtained by individuals with low-income level was higher when compared to individuals with a balanced income-expense level.

Adhering to a diet in addition to a gluten-free diet was associated with poorer psychosocial adjustment across all PAIS-SR domains. This finding suggests that the burden of dietary management increases considerably when individuals must follow multiple dietary restrictions.

In individuals with celiac disease, food choices are already limited, and the addition of further dietary rules may intensify feelings of restriction, social exclusion, and loss of autonomy. These individuals may experience greater difficulty in eating outside the home, in participating in social events, and in maintaining dietary flexibility. From a clinical perspective, this finding indicates that patients with multiple dietary requirements should be considered a particularly vulnerable group and may benefit from individualized dietetic counseling and psychosocial support.

The presence of another family member with celiac disease was associated with better healthcare orientation. This may indicate that shared illness experiences within the family facilitate access to practical knowledge, emotional support, and problem-solving strategies. Families with more than one affected member may develop greater awareness of gluten-free food preparation, cross-contamination prevention, and healthcare follow-up. Such shared experience may reduce uncertainty and promote more rapid adaptation to the disease. However, this finding should also be interpreted with caution, as family history may be associated with earlier diagnosis, greater disease awareness, or stronger engagement with patient associations.

Self-management findings also provide important insights into psychosocial adjustment. Higher self-stigma was associated with poorer psychosocial adjustment, whereas better coping with stigma, greater healthcare efficiency, and greater compliance with treatment were associated with better psychosocial adjustment. This pattern suggests that psychosocial adjustment is closely linked to how individuals perceive and manage the social meaning of their illness. Self-stigma may lead patients to conceal their condition, to avoid social eating situations, or to withdraw from social interactions, which may worsen psychological distress and impair adaptation. Conversely, effective coping skills, confidence in healthcare management, and adherence to treatment may strengthen resilience and support more positive adjustment to living with celiac disease. In line with this interpretation, Bektaş and Uysal²⁴ reported similar patterns among individuals with chronic illness, with relatively higher scores in healthcare efficiency and compliance with treatment than in self-stigma and coping with stigma. These findings suggest that strengthening disease-related self-management skills may be an important strategy for reducing stigma-related difficulties and improving psychosocial adjustment in individuals with chronic illnesses.

Gender-based differences in self-stigma may also be understood within the same social context. Greater visibility of dietary restrictions in family and social settings may increase female patients' feelings of difference or inadequacy, particularly when food preparation and caregiving responsibilities are culturally associated with women. However, this finding should be interpreted cautiously because the sample was predominantly female.

Marital status was also associated with self-management. Married patients had higher scores in the coping with stigma and compliance with treatment sub-dimensions than those of single patients. This result is consistent with the study by Bakır and Zengin,²⁵ which reported higher treatment compliance among married individuals with diabetes. This may suggest that spousal support not only improves emotional adjustment but also supports daily disease management behaviors, such as maintaining a gluten-free diet, seeking health information, and following treatment recommendations.

In terms of self-management, income level may influence patients' access to gluten-free foods, professional counseling, and reliable disease-related information. Therefore, better coping with stigma and greater healthcare efficiency observed among participants with more balanced incomes may reflect greater access to resources that support daily disease management. This interpretation is consistent with previous studies reporting better self-management outcomes among individuals with higher income levels.^{24,25}

No significant differences were found across education levels in the CISM sub-dimensions. This suggests that education level alone may not be sufficient to explain self-management in celiac disease. Disease-specific knowledge, access to professional counseling, social support, health literacy, and previous disease experience may be more influential than formal education level. In contrast, Uğuz and Doğan²⁶ reported that hemodialysis patients with a university education or higher had higher treatment compliance scores. This difference may be related to the distinct treatment requirements for hemodialysis and celiac disease, as celiac disease management depends heavily on daily food choices, environmental control, and long-term behavioral adaptations.

From a self-stigma perspective, adopting an additional diet may make the illness more visible in social contexts and increase the perception of being different from others. This finding may explain why participants with multiple dietary restrictions reported higher self-stigma scores.

Age was negatively correlated with self-stigma. This may indicate that as patients grow older, they become more familiar with their chronic illness and develop greater acceptance of disease-related limitations. Increased life experience may also help individuals manage social reactions more effectively and reduce the internalization of stigma. Similarly, longer disease duration was positively correlated with healthcare efficiency. This finding may reflect the gradual development of disease-specific knowledge, confidence in communicating with health professionals, and practical skills related to gluten-free diet management over time. Nevertheless, these associations should be interpreted as correlational and may also be influenced by unmeasured factors such as health literacy, family support, or access to follow-up care.

The relationship between the PAIS-SR total score and CISM sub-dimensions further supports the association between self-management and psychosocial adjustment. Higher self-stigma was associated with poorer psychosocial adjustment, suggesting that internalized stigma may negatively affect emotional well-being, social participation, and treatment adaptation. Patients who feel ashamed or different because of their illness may avoid social eating situations, conceal their diagnosis, or withdraw from social interactions. In contrast, coping with stigma, greater healthcare efficiency, and higher compliance with treatment were associated with better psychosocial adjustment. In support of the importance of treatment adherence in disease management, Enaud et al.²⁷ reported that adherence to a gluten-free diet and the duration of the diet were important factors associated with quality of life in patients with celiac disease. Their findings highlight the importance of improving dietary adherence and effective disease management to enhance the quality of life of individuals with celiac disease.²⁷ Taken together, these findings suggest that effective self-management, including treatment adherence and the ability to cope with disease-related challenges, may contribute to better psychosocial adjustment and overall well-being in individuals with celiac disease.

Overall, these findings suggest that self-management and psychosocial adjustment in celiac disease are shaped by interrelated social, economic, and psychological factors. Nursing interventions should therefore go beyond dietary education and include psychosocial counseling, stigma reduction strategies, family involvement, and support for patients with financial or dietary difficulties. Identifying vulnerable groups, such as women, single individuals, low-income patients, and those following additional diets, may help health professionals provide more individualized and patient-centered care. These findings are further supported by Zingone et al.,²⁸ who emphasized that psychological distress, anxiety, and reduced quality of life are common among individuals with celiac disease and may negatively affect disease management and long-term adaptation. Therefore, addressing psychosocial needs should be considered an essential component of comprehensive celiac disease care.²⁸

Study Limitations

This study has several limitations. First, participants were recruited from a single celiac association, which may have introduced selection bias. Individuals who are members of such associations may be more informed, motivated, and engaged in disease management than patients who are not affiliated with such organizations. Therefore, the findings may not fully represent all individuals with celiac disease. Second, the sample consisted predominantly of female participants (88.5%). This gender imbalance may have influenced the findings, particularly in gender-based comparisons, and may limit the generalizability of the results to male patients with celiac disease. Third, the study included participants from a limited geographical area; therefore, the findings may not be generalizable to all individuals with celiac disease in Türkiye. Because of the descriptive and correlational design of the study, causal relationships between self-management and psychosocial adjustment cannot be inferred.

CONCLUSION

This study concluded that the psychosocial adjustment and chronic illness self-management skills of individuals with celiac disease are affected by demographic and socioeconomic factors.

Recommendations

- **Psychosocial support and education:** Patients should be provided with special psychosocial support; interviews should be organized to prevent social isolation; and adjustment processes at home should be facilitated through patient education. In addition, age-appropriate educational programs should be organized.
- **Economic support and social awareness:** Food and financial aid programs should be developed to address the economic burden of a gluten-free diet; the sense of stigma should be reduced by increasing social awareness of celiac disease; and patients should be included in regular follow-up programs from the time of diagnosis.

MAIN POINTS

- Psychosocial adjustment was found to be poor in approximately one-third of the participants, indicating that celiac disease imposes not only a physical burden but also a significant psychosocial one.
- Factors such as female gender, low income, and adherence to additional dietary restrictions were associated with poorer

psychosocial adjustment. These findings highlight the need for targeted support programs for at-risk groups.

- Family and social support (particularly being married or having another family member with celiac disease) positively influence both treatment adherence and coping skills.
- Self-stigma is one of the key factors negatively affecting psychosocial adjustment. As stigma increases, adjustment decreases; conversely, as coping skills and treatment adherence improve, psychosocial adjustment strengthens.
- Self-management skills are strongly associated with psychosocial adjustment in adults with celiac disease.

ETHICS

Ethics Committee Approval: Ethical approval for the study was obtained from the Ankara Yıldırım Beyazıt University Medical Research Ethics Committee (approval number: 06/810, date: 01.07.2024).

Informed Consent: Written informed consent was obtained from all participants, and permission to use the scales was secured from the authors.

Footnotes

Authorship Contributions

Concept: İ.B., H.T., Design: İ.B., H.T., Data Collection and/or Processing: İ.B., Analysis and/or Interpretation: İ.B., H.T., Literature Search: İ.B., Writing: İ.B.

DISCLOSURES

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

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

Declaration on the Use of Artificial Intelligence (AI): Artificial intelligence has been used to assist with “text editing”.

REFERENCES

- Caio G, Volta U, Sapone A, Leffler DA, De Giorgio R, Catassi C, et al. Celiac disease: a comprehensive current review. *BMC Medicine*. 2019; 17(1): 142.
- Ben Houmich T, Admou B. Celiac disease: understandings in diagnostic, nutritional, and medicinal aspects. *Int J Immunopathol Pharmacology*. 2021; 35: 20587384211008709.
- Pinto-Sanchez MI, Silvester JA, Leibold B, Leffler DA, Anderson RP, Therrien A, et al. Society for the study of celiac disease position statement on gaps and opportunities in celiac disease. *Nat Rev Gastroenterol Hepatol*. 2021; 18(12): 875-84.
- Taşdelen Baş M, Çavuşoğlu H. Evaluation of daily lives of children and adolescents with celiac disease and nursing approaches. *Gastroenterol Nurs*. 2020; 43(6): E202-13.
- Centers for Disease Control and Prevention. (2024, May 15). Chronic diseases in America. Available from: <https://www.cdc.gov/chronic-disease/about/index.html>
- Ngai SSY, Cheung CK, Ng YH, Tang HY, Ngai HL, Wong KH. Development and validation of the chronic illness self-management (CISM) scale: data from a young patient sample in Hong Kong. *Child Youth Serv Rev*. 2020; 114: 105077.
- Schulman-Green D, Jaser SS, Park C, Whittemore R. A metasynthesis of factors affecting self-management of chronic illness. *J Adv Nurs*. 2016; 72(7): 1469-89.
- Vainauskienė V, Vaitkienė R. Enablers of patient knowledge empowerment for self-management of chronic disease: an integrative review. *Int J Environ Res Public Health*. 2021; 18(5): 2247.
- Dargahi S, Pireinaladin S, Ahmadboukani S, Bashirgonbadi S, Molaei B, Fathi D. The role of perceived burdensomeness, thwarted belongingness, and giving meaning to life in psychosocial adjustment: the mediating role of self-care in the elderly. *Iran J Psychiatry*. 2023; 18(2): 145-52.
- Togluk S, Çuhadar D. The effect of death anxiety on psychosocial adjustment in individuals with chronic obstructive pulmonary disease. *Indian J Palliat Care*. 2021; 27(3): 358-66.
- Öztürk YE, Yeşildal M, Arık Ö, Fidan Y. Turkish validity and reliability of the chronic disease self-management scale. *J Acad Value Stud*. 2021; 7(3): 375-81.
- Adaylar M. Kronik hastalığı olan bireylerin hastalığına tutum, adaptasyon, algı ve öz - bakım yönelimleri. İstanbul Üniversitesi, Doktora Tezi. 1995.
- Nunnally JC. *Psychometric theory*. 2nd ed. New York: McGraw-Hill; 1978.
- Tabachnick BG, Fidell LS. *Using multivariate statistics*. 6th ed. Boston: Pearson; 2013.
- Rose C, Law GU, Howard RA. The psychosocial experiences of adults diagnosed with coeliac disease: a qualitative evidence synthesis. *Qual Life Res*. 2024; 33(1): 1-16.
- Taghadossi M, Tajamoli Z, Aghajani M. Psychosocial adaptation and its related factors in cancer patients undergoing chemotherapy. *Biomed Res Ther*. 2017; 4(12): 1853-66.
- Öyke N. Behçet hastalarının hastalığına psikososyal uyumu. Düzce Üniversitesi, Yüksek Lisans Tezi. 2008.
- Çelik S, Kelleci M, Avcı D, Temel E. Tip 1 diyabetli genç yetişkinlerin hastalığına psikososyal uyumları ve stresle başa çıkma tarzları. *FN Hemşirelik Derg*. 2015; 23(2): 105-15.
- Akyüz M, Şahin Altun Ö. The effect of sleep quality on psychosocial adjustment in patients with heart failure: descriptive research. *Türkiye Klinikleri J Nurs Sci*. 2024; 16(3): 875-84.
- Türksen Ülkü S, Ünsar S. Determination of psychosocial adjustment of patients with heart failure. *Turkish Journal of Cardiovascular Nursing*. 2022; 13(32): 159-66.
- Senmar M, Hasannia E, Moenoddin A, Lotfi S, Hamed F, Habibi M, et al. Psychosocial adjustment to illness and its relationship with spiritual well-being in Iranian cancer patients. *Int J Chronic Dis*. 2020; 2020: 5742569.
- Vural Doğru B, Karadakovan A. Evaluation of the correlation between quality of life and psychosocial adaptation in elderly patients with heart failure. *Turk J Card Nur*. 2016; 7(13): 88-104.
- Çam O, Nehir S. Investigation of the relationship between anxiety and depression levels with psychosocial adaptation of patients who experienced myocardial infarction. *Ege Univ Hemşirelik Fak Derg*. 2011; 27(2): 47-59.
- Bektaş G, Uysal N. Determination of self-management and health seeking behaviors in individuals with chronic disease: descriptive study. *Türkiye Klinikleri J Nurs Sci*. 2024; 16(1): 207-18.
- Bakır GG, Zengin N. Investigation of self-management and affecting factors in individuals with diabetes. *Sağlık ve Yaşam Bilimleri Derg*. 2023; 5(1): 9-17.
- Uğuz Ö, Doğan S. The relationship between chronic disease self-management and well-being in hemodialysis patients. *J Nephrol Nurs*. 2023; 18(2): 87-98.
- Enaud R, Tetard C, Dupuis R, Laharie D, Lamireau T, Zerbib F, et al. Compliance with gluten free diet is associated with better quality of Life in celiac disease. *Nutrients*. 2022; 14(6): 1210.
- Zingone F, Swift GL, Card TR, Sanders DS, Ludvigsson JF, Bai JC. Psychological morbidity of celiac disease: a review of the literature. *United European Gastroenterol J*. 2015; 3(2): 136-45.

Compliance with Tobacco Control Law in Hospitals of Nicosia: An Observational Study

Özen Aşut¹, Songül Acar Vaizoğlu², Şanda Çalı¹, Ahmed Elamin Ahmed Elamin³, Mohamed Ebied³, Enver Abdullaiev³, Karam Chaddad³, Muhnd Muhammad Algawi³, Shania Sheka³, Abdullatif Adnan İbrahim³, Tuqa Mohammad Aldowery³, Gülşah Bahadır³, Mehmet Utku Durum³, Alper Çeliker³

¹Department of Public Health, Near East University Faculty of Medicine, Nicosia, Cyprus

²Department of Public Health, Girne University Faculty of Medicine, Nicosia, Cyprus

³MD Freelance, Nicosia, Cyprus

Abstract

BACKGROUND/AIMS: Smoking is a global health issue causing millions of deaths each year. The World Health Organisation created the Framework Convention of Tobacco Control and launched the MPOWER initiative in 2008 to reduce this death burden. Studies conducted in developed countries show that strict tobacco control laws help reduce tobacco use. In the Turkish Republic of Northern Cyprus (TRNC), studies show a high prevalence of tobacco use, but there are no reports on the implementation of laws and adherence to them. This research aims to evaluate compliance with the tobacco control law in three hospitals in Northern Cyprus.

MATERIALS AND METHODS: This observational study was conducted in August 2025. Observers used a data collection form to assess compliance. SPSS version 31.0 (IBM Inc., Armonk, NY, USA) was used for data analysis.

RESULTS: A total of 113 departments were observed across three hospitals. Indoor no smoking compliance was high, whereas overall compliance with “no smoking” signs was low (37.4%). The highest compliance was found in the state hospital, particularly in signage placement, while private hospitals showed gaps, such as insufficient outdoor signage and the presence of smoking facilities. No enforcement officers or cessation resources were observed. Staff smoking was observed in all hospitals. A significant difference in compliance between the state hospital and the private hospitals was observed, with the state hospital having better compliance (p-value <0.001).

CONCLUSION: Compliance with the Northern Cyprus tobacco control law varied by hospital type. Major shortcomings were found in signage, enforcement, and cessation support. No hospital had trained personnel to handle violations, and hospital staff were observed smoking.

Keywords: Tobacco control law, smoke-free policy, hospital compliance, enforcement, Northern Cyprus, TRNC

To cite this article: Aşut Ö, Acar Vaizoğlu S, Çalı Ş, Ahmed Elamin AE, Ebied M, Abdullaiev E, et al. Compliance with tobacco control law in hospitals of Nicosia: an observational study. Cyprus J Med Sci. 2026;11(5):401-408

ORCID IDs of the authors: Ö.A. 0000-0002-9604-4037; S.A.V. 0000-0001-6962-4194; Ş.Ç. 0000-0001-9929-2637; A.E.A.E. 0009-0002-1598-7111; M.E. 0009-0006-1502-5034; E.A. 0009-0007-2924-6708; K.C. 0009-0008-5509-4688; M.M.A. 0009-0007-8791-9552; S.S. 0009-0009-6749-8660; A.A.İ. 0009-0000-0195-5800; T.M.A. 0009-0009-9132-9648; G.B. 0009-0008-5809-7470; M.U.D. 0009-0000-5709-013X; A.Ç. 0009-0006-2251-0919.



Corresponding author: Ahmed Elamin Ahmed Elamin

E-mail: drahmedelamin2001@gmail.com

ORCID ID: orcid.org/0009-0002-1598-7111

Received: 12.02.2026

Accepted: 03.08.2026

Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.

This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

INTRODUCTION

The World Health Organisation (WHO) created the Framework Convention of Tobacco Control (FCTC) and established the MPOWER initiative in 2008.¹ This was due to the high death burden caused by smoking, which accounted for over eight million deaths annually.²

These initiatives have encouraged many countries to develop stronger tobacco control legislation; currently, 6 billion people (61% of the global population) are covered by at least one tobacco control measure aimed at reducing tobacco use. This shows a notable progress when compared to 2007, where only 1 billion (15%) had similar protection.³

Smoking rates in many developed countries have fallen sharply in the last 3 decades. In the United States, prevalence dropped from 25.3% in 1990 to 19.8% in 2022,^{4,5} while in the European Union, prevalence decreased from 32% in 2006 to 18.4% in 2022.^{6,7} This decrease is mainly due 2009 Family Smoking Prevention and Tobacco Control Act and Directive 2014/40/EU.^{8,9} However, evidence depicts that legislation alone is not sufficient; real progress depends on consistent enforcement, active monitoring, and institutional involvement.^{10,11}

A comprehensive set of smoking-control policies has been reviewed in the United States and other countries and has been shown to be effective in preventing smoking. Integrated policies, particularly strong smoking restrictions, cigarette taxation, and clean air laws, may reduce the prevalence of tobacco consumption.¹²

Real-world implementation of tobacco laws in smoke-free places, such as hospitals, can be challenging. For example, in Türkiye, Armenia, and Australia, concerns such as staff smoking on duty, insufficient signage placement, lack of outdoor enforcement, and limited cessation support materials were observed.¹³⁻¹⁵ However, Amsterdam and Catalonia showed the opposite, where enforcement was consistent and combined with active communication and cessation support.^{16,17} A report in Spain indicated that patients viewed smoke-free compliance in hospitals to be high, highlighting the need for hospitals to maintain the expected level of compliance.¹⁸

Regionally, data show unfavorable findings on smoking prevalence in healthcare settings. A study based at Nicosia General Hospital in Cyprus showed a high prevalence of smokers in the hospital (28.2%), slightly higher than the general population prevalence in Southern Cyprus.¹⁹ Greece demonstrated even poorer performance in this regard. Smoking practices among medical professionals and nursing staff in a university hospital in Greece indicated a concerning smoking prevalence of 57.8% among nurses and 34.5% among physicians, according to research conducted in 2009.²⁰

In Northern Cyprus, evidence is limited, but available studies are alarming. Two surveys of healthcare professionals in a military hospital and medical students have shown smoking prevalence of 53.4% and 33.7%, respectively.^{21,22} Another study that assessed medical students from four faculties in Türkiye and Northern Cyprus reported that 19.9% were current smokers and showed high passive smoke exposure, and poor enforcement of bans.²³ A cross-sectional survey of visitors to a primary healthcare center in Nicosia showed that 40.7% of the participants were smokers. Notably, 64% of respondents in this survey found local tobacco control efforts to be lacking.²⁴

The Turkish Republic of Northern Cyprus (TRNC) enacted its first comprehensive tobacco control legislation in 2008, banning smoking in indoor public places, restricting marketing and sales, and mandating health warnings on packages, and "no smoking" signs.²⁵

Although the 2008 legislation complies with several of the WHO FCTC's most significant requirements, implementation and enforcement of the legislation in the TRNC are commonly underreported. In the Eastern Mediterranean, the WHO states that most countries have smoke-free laws, but they are being enforced and implemented poorly.²⁶⁻²⁸

Findings from similar studies in other regions point towards weak compliance: a study at a Bangladeshi university found weak compliance with tobacco control legislation, as indicated by weak signages and weak anti-tobacco messages.²⁹ In addition, findings from observations in Delhi government hospitals showed a composite rate of only 45%, in addition to the existence of active smoking and tobacco shops on or within hospital grounds.³⁰

To our knowledge, no studies have been conducted to evaluate tobacco control compliance in hospitals in the TRNC. This is an important gap in the literature, particularly given the elevated rates of smoking prevalence identified among healthcare workers, students and the general population.²¹⁻²⁴ As no published compliance data are available, the present study attempted to establish compliance with smoke-free policies in private, university, and state hospitals in Northern Cyprus through systematic on-site surveys using a standardized checklist, thus providing an objective measure of actual compliance.

Objective

The general objective of our study is to observe and assess adherence to the TRNC tobacco control law in 3 hospital settings and to determine whether compliance differs within the same hospital or across hospitals.

Specific Aims:

1. Assess compliance within different departments.
2. Identify any violations of the law within different individual department service areas (e.g., patient ward, polyclinic, waiting rooms, and dining areas)
3. Evaluate compliance in shared service areas (e.g., dining areas and outdoor areas)
4. Assess hospital-specific checks, such as smoking signage, staff smoking, tobacco sales, staff trained to handle violations, and availability of smoking cessation resources.

MATERIALS AND METHODS

The study was conducted in three hospitals in Nicosia, Northern Cyprus. These hospitals include a state hospital, a university-affiliated private hospital, and a private non-university hospital. Field observations were done between the 6th and 12th of August 2025. A pre-test was performed on the 5th of August 2025 in a hospital other than the study hospitals.

This descriptive, observational study aimed to assess the hospitals' compliance in Nicosia, TRNC, with the Tobacco Control Law of Northern Cyprus. This study employed a data collection form developed by the researchers following a literature review.³¹

Observation Strategy

The observers were trained adequately before data collection to avoid inconsistencies. They served only as data collectors for research purposes and did not have any legal or administrative authority to enforce tobacco control laws or to intervene when violations were observed.

The observer team consisted of eleven medical students; no inspectors were appointed by the public health authority, nor were any participants from the general public included. All observations were performed directly by the trained observer team using a standardized checklist.

Observers were divided according to the number of departments in each hospital, and then gathered data using a checklist for their compliance assessment. Their observations were made during daytime hours, and their recorded data were based on observations during one week.

The departments covered were general medicine, surgical, obstetric, and laboratory units. All of their related subspecialties were also evaluated. Administrative units, pharmacies, and dental clinics were also covered.

The assessment aimed to evaluate the presence and visibility of “no smoking” signage, including whether the signs displayed only the smoking prohibition or included information on financial penalties. The assessment covered all tobacco products regulated under the TRNC tobacco control law, including electronic cigarettes and heated tobacco products. Advertisements or sales of tobacco products, staff adherence to the smoke-free law, and compliance in general areas, such as outdoor areas and around hospital entrances, were also assessed. Visual estimation was the accepted method for determining tobacco sales within 100 meters of the hospitals.

Study Variables

The dependent variable is compliance with the tobacco control law in hospitals, measured by indicators such as the presence or absence of “no smoking” signage in wards, waiting and dining areas, and staff behaviours.

Independent variables included hospital category (public, private, or university), facility type, department type, administrative measures such as trained personnel for violations, and tobacco sales.

Data Collection Form (Appendix 1)

The data collection form was divided into 4 sections: compliance in specific service locations inside the hospital (section 1), compliance in service areas outdoors (section 2), compliance in general areas (section 3), and compliance with hospital-specific checks (section 4). Sections had 34 items in total, and were answered using a “yes or no” format.

Section 1 consisted of items assessing specific service areas, including wards, polyclinics, waiting areas and dining areas for the presence of “no smoking” signs and whether active smoking is observed (none, one person smoking, or multiple smoking).

Section 2 had assessment items regarding the general service areas, including cafeterias and outdoor areas (open and semi-open spaces) with the goal of observing the presence of selling cigarettes to minors, “no sales to minors” sign, and if a signage was positioned within an eight-meter boundary from the openings and entrances.

Section 3 assesses compliance in general areas by checking if any smoking litter and ashtrays were noticed, “no smoking” signs were correctly placed, and whether or not designated smoking areas or tobacco sales were present.

Section 4 monitored specific hospital checks, ensuring that the smoking ban covered hospital gates, departments, wards, and waiting areas. Staff compliance with tobacco-use restrictions and the presence of security personnel to handle violations were observed. The presence of tobacco sales within 100 meters of hospital boundaries and the availability of smoking cessation resources, such as pamphlets, were also inspected. Visual estimation was the accepted method for determining tobacco sales within 100 meters of the hospitals.

Pre-Test

The pretest was conducted on 5 August 2025 at three hospitals, which were not included in the main study. Eleven forms were completed at these three hospitals to test the clarity and practicality. According to the pre-test, no revisions were needed in the data collection form.

Ethical Considerations

The study is based on observations and no humans were involved in the study. Thus, it was concluded that is ethical approval was not required.

Researchers

This research was conducted by eleven sixth-year medical students and three academic members of the Near East University faculty of medicine department of public health.

Data Analysis

The statistical analysis was performed using IBM SPSS version 31.0 (IBM Inc., Armonk, NY, USA) was used for data analysis. Descriptive statistics (frequencies, percentages) were generated. Comparative analyses between hospital categories (state, university-private, and non-university-private) and service areas were conducted using chi-square tests. The standard alpha level ($\alpha = 0.05$) was accepted as the threshold for statistical significance. A p-value < 0.05 was considered statistically significant. Hospitals were also grouped to compare the state hospital with the other two private hospitals.

TRNC Tobacco Control Law (Compliance Conditions)

The TRNC Tobacco Control Law sets out the requirements for compliance.

Purpose and scope: The purpose of this law is to protect individuals, including children and youth, from the harmful effects of tobacco products. The law regulates the production, importation, sale, advertising, and use of tobacco products in the TRNC. The law applies to all public and private spaces within the TRNC.

Packaging and warning labels: Tobacco product packaging must include health warnings and product content information. The packages' visible sides must display warnings such as “smoking kills” and indicate the levels of tar, nicotine, and carbon monoxide.

Smoking bans and designated areas: Smoking is banned in the indoor areas of all public buildings, educational and healthcare institutions, public vehicles, cafes, and restaurants. Exceptions are made for designated smoking areas in certain places such as prisons, care homes, and certain bars. These areas should be well-ventilated

and inaccessible to minors. Such smoking areas must not exceed 50% of the total area.

Sales restrictions: Selling tobacco products to individuals under the age of 18 is prohibited. Identity verification is mandatory. Sales through vending machines are banned. The production and sales of products that resemble tobacco, such as candy and toys, are also prohibited. Furthermore, sales locations must be arranged out of children's reach.

Advertising, promotion, and sponsorship ban: All forms of direct and indirect tobacco advertising, promotions, and sponsorships are banned. Companies or organizations cannot sponsor any event that uses tobacco brand names or symbols.

Enforcement and penalties: Inspectors appointed by the Ministry of Health are authorized to supervise the implementation. They may issue administrative fines and stop unlawful practices. Penalties are calculated as multiples of the minimum wage and may include imprisonment for repeated violations.

Education and awareness: Television and radio stations are required to broadcast at least 90 minutes of public health programming per month to raise awareness about tobacco use. Warning signs must be clearly displayed in all areas where smoking is prohibited.

RESULTS

One hundred and thirteen data points were designated across the three hospital categories. The study encompassed a variety of hospital sections to ensure sufficient data collection. Table 1 shows the distribution of the hospitals as state, non-university private, and private university, and also the distribution of the departments according to medical disciplines.

According to the analysis, the state hospital constituted the largest share of observed facilities (39.8%, $n=45$). It is closely followed by the university private hospital (35.4%, $n=40$). The non-university private hospital accounted for the smallest proportion (24.8%, $n=28$).

The internal medicine departments were identified as the primary sites for data collection, as over half of the observations (54.9%, $n=62$) occurred in this group. Surgical departments accounted for 28.3% ($n=32$) of the total observations, while laboratories accounted for only 8.8% ($n=10$). Other departments, including administrative units,

audiology, pharmacy, nutrition and dietetics, and oral and dental clinics, accounted for a small proportion (8%, $n=9$).

Table 2 shows the status of compliance with control measures at specific service locations within the hospitals.

Table 2 outlines observations in different hospital areas concerning indoor smoking and the presence of no smoking signage. Very high compliance with the no smoking policy was observed in all service areas of the hospitals. Dining rooms had the highest compliance (96.5%, $n=109$); waiting rooms had the lowest (92.7%, $n=102$). One-person smoking was reported in all locations, most commonly in waiting rooms (6.4%, $n=7$). Higher use was not observed, except in waiting and dining rooms (0.9%, $n=1$ each). Most locations did not have clearly visible no smoking signs, with dining rooms showing the worst compliance: only 31.9% ($n=36$) had relevant signs. Overall, 37.4% of these areas had "no smoking" signage.

Table 3 displays tobacco-control compliance in hospital canteens, outdoor areas, and common areas.

Table 3 presents the evaluation of compliance across three hospitals in Nicosia: a non-university private hospital, a university private hospital, and a state hospital. The table is organized into three domains: service area and outdoor compliance, general area compliance, and hospital-specific checks.

The state hospital showed high compliance by correctly placing "no smoking" signage within 8 meters of enclosed, unenclosed, and outdoor areas. The university's private hospital failed to display "no smoking" signage, reflecting low overall compliance. The non-university private hospital complied with signage placement near enclosed areas, but failed to display signage in unenclosed and general outdoor areas. None of the hospitals had "no sales to minors" signage, and only the university's private hospital lacked an enforced ban on tobacco sales in the hospital canteen.

The state and university private hospitals showed strong compliance across general areas. Both displayed legally sized "no smoking" signs at all entrances. Tobacco sales were prohibited, and no evidence of smoking or designated smoking areas was found in the state and non-university private hospitals. The university hospital allowed tobacco sales at several points on campus.

Although the non-university private hospital had no evidence of tobacco sales or advertising, it maintained a designated smoking area, failed to display signage at all entrances, and its signage did not meet legal size requirements, which constitutes a clear violation of policy.

All three hospitals lacked an enforcement sign identifying the personnel responsible for any violations.

Compliance with the hospital-specific checks was variable. A ban on smoking inside facilities was upheld across all hospital types. The state and university private hospital constantly displays "no smoking zone" signage at all hospital gates and throughout the campus.

Key limitations were noted in the enforcement of the law. Neither hospital had trained security to handle violations. Likewise, all hospitals lacked smoking cessation resources, anti-smoking pamphlets, a cessation helpline, and counselling services. Smoking among hospital staff was also a common issue across all three hospitals.

Table 1. Distribution of hospitals and departments (Nicosia, August 2025) $n=113$

Departments per hospital ($n=113$)	n	%
State	45	39.8
Private university	40	35.4
Non-university private	28	24.8
Distribution of the departments as to medical disciplines ($n=113$)		
Internal medical departments	62	54.9
Surgical departments	32	28.3
Laboratories	10	8.8
Others*	9	8.0
*Others include administrative units, audiology, pharmacy, nutrition and dietetics, and oral and dental clinics.		

Table 2. Compliance in service locations within the hospitals in regard to indoor smokers and no smoking signage (Nicosia, August 2025) n=113

Location	No smoker observed		One smoker observed		More than one smoker observed		“No smoking” signage exists	
	n	%	n	%	n	%	Yes	
Patient wards (n=108)	104	96.3	4	3.7	0	0	44	40.7
Polyclinic (n=108)	104	96.3	4	3.7	0	0	43	39.8
Waiting room (n=110)	102	92.7	7	6.4	1	0.9	41	37.3
Dining room (n=113)	109	96.5	3	2.7	1	0.9	36	31.9

Table 3. Compliance in canteens, outdoors, and general common areas (Nicosia, August 2025) n=3

Feature	Hospital		State		University private		Non-university private		Score	
			Yes	No	Yes	No	Yes	No	Y	N
Compliance in service areas and outdoors										
No tobacco sales in the hospital canteen			+			+	+		1	-
No sales to minors within the canteen			+			+	+		1	-
“No sales to minors” signage within the canteen				+		+		+	1	-
“No smoking” sign within 8 meters of enclosed areas			+			+	+		1	-
“No smoking outdoors” signage			+			+		+	1	-
“No smoking” sign within 8 meters from open unenclosed areas			+			+		+	1	-
“No smoking” signage outdoors			+		+		+		1	-
General area compliance										
Prohibited ash cans or ashtrays				+	+			+	1	-
Enforcement signage (responsible person displayed)				+		+		+	1	
“No smoking” signage at all entrances			+		+			+	1	-
Signage meets legal size			+		+			+	1	-
No evidence of smoking			+		+		+		1	-
No “designated smoking area”			+		+			+	1	-
No tobacco sales within the hospital premises			+			+	+		1	-
No advertisement of tobacco			+		+		+		1	-
Awareness materials displayed				+		+		+	1	-
Hospital-specific checks										
“No smoking zone” signage at all gates and entrances			+		+			+	1	-
No smoking inside hospital facilities*			+		+		+		1	-
No smoking by hospital staff				+		+		+	1	-
Trained security to handle violations				+		+		+	1	-
No tobacco sales within 100 meters			+			+		+	1	-
No street sales			+		+		+		1	-
Smoking cessation resources				+		+		+	1	-
Anti-smoking pamphlets				+		+		+	1	-
Smoking cessation helpline				+		+		+	1	-
Counselling services				+		+		+	1	-
Total score and %**			16	61.5%	10	38.5%	9	34.6%	26 (100)	

*Wards, outpatient clinic, waiting areas, intensive care unit, etc, **Row percentages.

Regarding boundary control, the state hospital was fully enforcing the ban on tobacco sales within 100 meters of the premises, while neither the university-affiliated private hospital nor the non-university private hospital was enforcing the rule. Furthermore, all hospitals complied with the prohibition on street sales outside their premises.

Moreover, a scoring system was used to evaluate the level of compliance among the three hospitals; a total of 26 compliance measures were assessed. Each hospital is awarded one point of adherence, “yes,” and zero for non-adherence, “no”. The state hospital demonstrated the highest level of compliance, achieving 16/26 points (61.5%). The state hospital demonstrates a moderate level of enforcement of tobacco

control measures. On the other hand, the university private hospital scored 10/26 points (38.5%), while the non-university private hospital scored 9/26 (34.6%), both demonstrating lower compliance.

The results shown in Table 2 demonstrate that the state hospital had the highest overall compliance, especially in signage and perimeter control. The university’s private hospital met most indoor requirements, but lacked outdoor signage and permitted tobacco sales inside. The non-university private hospital demonstrated the lowest level of compliance, most notably in signage and policy enforcement. The absence of enforcement personnel and of cessation support services, together with staff smoking on duty, are major limitations.

No smoking signage was significantly more common in the state hospital compared to the two private hospitals across all areas: patient wards 80.0% (n=36) vs. 12.7% (n=8), polyclinics 95.6% (n=43) vs. 0, waiting rooms 91.1% (n=41) vs. 0, and dining rooms 77.8% (n=35) vs. 1.5% (n=1). The state hospital demonstrated markedly higher compliance, and the differences were statistically significant, as shown in Table 4 and Figure 1.

Compliance with the requirement for “no smoking zone” signage at all entrances was significantly higher in the state hospital, 77.8% than in the 2 private hospitals, 44.1% (p<0.001). Overall, 57.5% (n=65) of all entrances had “no smoking zone” signage (Table 5).

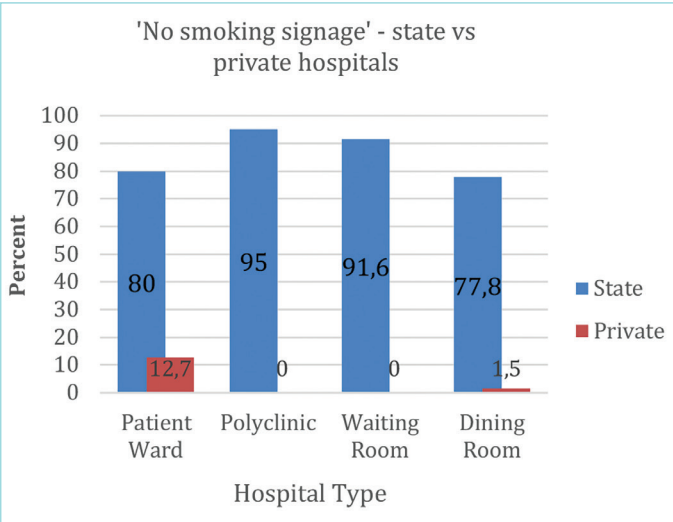


Figure 1. Presence of “no smoking” signage in the state hospital versus the two private hospitals, by area.

DISCUSSION

Research focusing on compliance with tobacco control in the TRNC is limited. This study is the first to evaluate compliance levels among hospitals in the TRNC. Global evidence shows that the implementation of indoor tobacco control policies results in a decline in violations, even in settings with weak enforcement as shown in the study from Spain.¹⁸ This matches the findings of our study, as compliance within indoor hospital areas exceeded 92%, which may explain why active smoking was rarely observed there.

The results of our study indicated inadequate “no smoking” signage in general areas, outdoor spaces, and departments, with only 37% of departments meeting the standard. Public hospitals demonstrated more consistent compliance than private hospitals, which lacked proper signage. This may be due to stronger regulatory oversight of public hospitals by the government. More research is needed to provide stronger evidence for this relationship.

Supporting this, a study in Delhi found higher compliance in tertiary hospitals that are under stronger government control.³⁰

The study found that tobacco sales occurred within 100 meters of private hospitals, while no sales were observed near the state hospital. Similarly, tobacco sales near healthcare facilities were observed in countries such as Türkiye, Armenia, and Bangladesh.^{13,14,29}

All three hospitals lacked enforcement mechanisms and cessation support. For example, none of the hospitals had trained personnel or signage indicating the personnel responsible for handling violations. No cessation support was available, in contrast to Catalonia and Amsterdam. In these regions, cessation support services lead to smoke-free policy implementation, higher compliance, and better health outcomes.^{16,17}

However, Australia showed that even with the presence of cessation support, there can still be low compliance.¹⁵ These findings indicate that cessation support services alone are insufficient and must be combined with the active enforcement of other aspects of tobacco control to be effective.

Smoking was observed among staff in all hospitals, raising concerns about the validity of enforcement and the role-model image of healthcare workers. Studies in the Eastern Mediterranean region and Cyprus showed that smoking ranged from 20% to 50% depending on the country.^{19,21-24,28} This can explain the violations observed within the hospital and indicates that cessation programs for healthcare workers should be mandatory.

Table 4. Presence of “no smoking” signage in the state versus the two private hospitals (Nicosia, August 2025) n=113													
Hospital type	Private hospitals				State hospital				Total				X ² p
	Yes		No		Yes		No		Yes		No		
	n	%	n	%	n	%	n	%	n	%	n	%	
Patient ward	8	12.7	55	87.3	36	80.0	9	20.0	44	40.7	64	59.3	49.25 <0.001
Polyclinic	0	-	64	100	43	95.6	2	4.4	43	39.8	65	60.2	100.03 <0.001
Waiting room	0	-	65	100	41	91.1	4	8.9	41	37.3	69	62.7	94.41 <0.001
Dining room	1	1.5	67	98.5	35	77.8	10	22.2	36	31.9	77	68.1	72.63 <0.001

Table 5. The presence of “no smoking zone” signage at all entrances			
Hospital type	“No smoking zone” sign		
		Yes	No
Private hospitals	n	30	38
	%	44.1	55.9
State hospital	n	35	10
	%	77.8	22.2
Total	n	65	48
	%	57.5	42.5
p-value	<0.001		

Studies from middle-income countries such as Türkiye, and Armenia found similar challenges as compliance was poor despite strong legislation.^{13,14} By contrast, studies from Amsterdam and Catalonia show that improvement can be achieved through institutional responsibility, active communication, and integrated cessation support.^{16,17}

Finally, the main findings highlight important public health issues. The insufficiency of visible signage and enforcement processes reduces the validity of the law, leading to partial compliance. Furthermore, the complete absence of cessation support materials in hospitals represents a missed public health opportunity, as hospitals should be places that encourage smoking cessation. Staff smoking weakens enforcement of the law since they are public role models for a healthy lifestyle. Therefore, the education and responsibility of healthcare workers are necessary to maintain support and to strengthen compliance.

Study Limitations

Our data covered only three hospitals and therefore may not be representative of all hospitals. It implies that we cannot conclude that our results are generalizable to all settings. Nevertheless, the research provides us with valuable information to understand the level of adherence to tobacco control policies, identifies areas for improvement, and guides future research.

CONCLUSION

Compliance is greater in state hospitals than in private hospitals. Observed problems included improper signage placement, staff smoking, a lack of trained staff to handle violations and issue warnings, and the absence of cessation support. The above observations imply that hospitals must improve staff training and implement a strict smoke-free policy.

Recommendations

To increase compliance with tobacco control in TRNC hospitals, we recommend the following steps:

- “No smoking” signs should be placed conspicuously at all major entrances.
- A “no smoking” policy should be implemented in outdoor areas, and the signs should be placed at strategic locations.
- Trained personnel should be present to address violations.

- Hospitals need to provide cessation support and guidelines, which need to be led by doctors and supported by other health professionals.
- Broader studies should be conducted to evaluate the effectiveness of the policy across all hospitals.

MAIN POINTS

- This is the first study to systematically examine hospital compliance with the tobacco control law within the Turkish Republic of Northern Cyprus.
- Indoor compliance is high across all hospitals.
- The state hospital exhibits the highest level of compliance overall.
- None of the hospitals had staff trained to deal with violations, and staff were observed smoking.
- Tobacco cessation initiatives or support were not available at any hospital.

ETHICS

Ethics Committee Approval: Not required, as no human participants were involved and no personal or patient-identifying data were collected.

Informed Consent: Not required, as no human participants were involved and no personal or patient-identifying data were collected.

Acknowledgements

The authors thank the medical students and faculty staff who contributed to data collection.

Footnotes

Authorship Contributions

Concept: Ö.A., S.A.V., Ş.Ç., Design: Ö.A., S.A.V., Ş.Ç., Data Collection and/or Processing: Ö.A., S.A.V., Ş.Ç., A.E.A.E., M.E., E.A., K.C., M.M.A., S.S., A.A.İ., T.M.A., G.B., M.U.D., A.Ç., Analysis and/or Interpretation: Ö.A., S.A.V., Ş.Ç., A.E.A.E., M.E., E.A., K.C., M.M.A., S.S., A.A.İ., T.M.A., G.B., M.U.D., A.Ç., Literature Search: Ö.A., S.A.V., Ş.Ç., A.E.A.E., M.E., E.A., K.C., M.M.A., S.S., A.A.İ., T.M.A., G.B., M.U.D., A.Ç., Writing: Ö.A., S.A.V., Ş.Ç., A.E.A.E., M.E., E.A., K.C., M.M.A., S.S., A.A.İ., T.M.A., G.B., M.U.D., A.Ç.

Conflict of Interest: Three authors of this article, Özen Aşut, Songül Acar Vaizoğlu and Şanda Çalı, are members of the Editorial Board of the Cyprus Journal of Medical Sciences. However, they were not involved in any stage of the editorial decision of the manuscript. The editors who evaluated this manuscript are from different institutions. The other authors declared no conflict of interest.

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

Declaration on the Use of Artificial Intelligence (AI): No artificial intelligence tools were used in the preparation of this manuscript.

Appendix 1: <https://d2v96fxpocvxx.cloudfront.net/beb8919b-f013-4ea1-b1c8-40332e840fe1/content-images/6f00355e-7e5b-47a5-8b8a-c4f391bd1c38.pdf>

REFERENCES

- World Health Organization. WHO report on the global tobacco epidemic, 2019: offer help to quit tobacco use. Geneva: World Health Organization; 2019. Available from: <https://www.who.int/publications/i/item/9789241516204>
- World Health Organization. WHO report on the global tobacco epidemic, 2023: Protect people from tobacco smoke. Geneva: World Health Organization; 2023. Available from: <https://www.who.int/publications/i/item/9789240077164>
- World Health Organization. WHO report on the global tobacco epidemic, 2025: warning about the dangers of tobacco. Executive summary [Internet]. Geneva: World Health Organization; 2025. Available from: <https://escholarship.org/uc/item/0cj684r8>
- National Center for Health Statistics. Table SmokSex. Current cigarette smoking among adults aged 18 and over, by sex, race, and age: United States, selected years 1965-2019. In: Health, United States, 2020-2021 [Internet]. Hyattsville, MD: National Center for Health Statistics; 2021 [cited 2025 Aug 21]. Available from: <https://www.cdc.gov/nchs/data/hsu/2020-2021/SmokSex.pdf>
- Centres for Disease Control and Prevention. Tobacco product use among adults—United States, 2022. Atlanta, GA: CDC; 2024.
- Bogdanovica I, Godfrey F, McNeill A, Britton J. Smoking prevalence in the European Union: a comparison of national and transnational prevalence survey methods and results. *Tob Control*. 2011; 20(1): e4.
- OECD/European Commission. Health at a glance: Europe 2024: state of health in the EU cycle. Paris: OECD Publishing; 2024. Available from: <https://doi.org/10.1787/b3704e14-en>
- Family Smoking Prevention and Tobacco Control Act. Pub L No. 111-31, 123 Stat 1776 (Jun 22, 2009) [Internet]. Washington (DC): US Government Publishing Office; 2009 [cited 2026 Jan 12]. Available from: <https://www.govinfo.gov/content/pkg/PLAW-111publ31/pdf/PLAW-111publ31.pdf>
- European Parliament and Council of the European Union. Directive 2014/40/EU on the approximation of the laws, regulations and administrative provisions of the Member States concerning the manufacture, presentation and sale of tobacco and related products. Official Journal of the European Union. 2014; 127: 1-38. Available from: <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32014L0040>
- Institute for Global Tobacco Control. Assessing compliance with tobacco advertising, promotion, and sponsorship (TAPS) bans: a “how-to” guide for conducting compliance studies of point-of-sale advertising and product display; outdoor advertising; and product packaging. Baltimore (MD): Johns Hopkins Bloomberg School of Public Health; 2013. Available from: https://theunion.org/sites/default/files/2020-08/TAPS_Compliance_eng.pdf
- Wynne O, Guillaumier A, Twyman L, McCrabb S, Denham AMJ, Paul C, et al. Signs, fines and compliance officers: a systematic review of strategies for enforcing smoke-free policy. *Int J Environ Res Public Health*. 2018; 15(7): 1386.
- Levy DT, Chaloupka F, Gitchell J. The effects of tobacco control policies on smoking rates: a tobacco control scorecard. *J Public Health Manag Pract*. 2004; 10(4): 338-53.
- Navas-Acien A, Çarkoğlu A, Ergör G, Hayran M, Ergüder T, Kaplan B, et al. Compliance with smoke-free legislation within public buildings: a cross-sectional study in Turkey. *Bull World Health Organ*. 2016; 94(2): 92-102.
- Movsisyan NK, Petrosyan V, Harutyunyan A, Petrosyan D, Stillman F. Clearing the air: improving smoke-free policy compliance at the national oncology hospital in Armenia. *BMC Cancer*. 2014; 14: 943.
- McCrabb S, Baker AL, Attia J, Balogh ZJ, Lott N, Palazzi K, et al. Hospital smoke-free policy: compliance, enforcement, and practices. a staff survey in two large public hospitals in Australia. *Int J Environ Res Public Health*. 2017; 14(11): 1358.
- Garritsen HH, Vermeulen JM, Rozema AD, van Lonkhuijzen LRCW, Kunst AE. Impact of a smoke-free policy on smoking prevalence on hospital grounds: a before-after study. *Tob Prev Cessat*. 2022; 8: 20.
- Martínez C, Fu M, Martínez-Sánchez JM, Ballbè M, Puig M, García M, et al. Tobacco control policies in hospitals before and after the implementation of a national smoking ban in Catalonia, Spain. *BMC Public Health*. 2009; 9: 160.
- Fu M, Castellano Y, Feliu A, Saura J, Estrada J, Galimany-Masclans J, et al. Compliance with the smoke-free policy in hospitals in Spain: the patients' perspective. *Eur J Cancer Prev*. 2023; 32(1): 81-8.
- Zinonos S, Zachariadou T, Zannetos S, Panayiotou AG, Georgiou A. Smoking prevalence and associated risk factors among healthcare professionals in Nicosia general hospital, Cyprus: a cross-sectional study. *Tob Induc Dis*. 2016; 14: 14.
- Vardavas CI, Bouloukaki I, Linardakis MK, Tzilepi P, Tzanakis N, Kafatos AG. Smoke-free hospitals in Greece: personnel perceptions, compliance and smoking habit. *Tob Induced Dis*. 2009; 5: 8.
- Günay E, Simşek Z, Kutucularoğlu G, Metinyurt G. Kuzey Kıbrıs Türk Cumhuriyeti askeri sağlık personelinde sigara içme sıklığı, nikotin bağımlılığı ve ucuz sigara satışının etkisi [Smoking prevalence, nicotine dependence and effects of low cost cigarette sale among military healthcare personal in Turkish Republic of Northern Cyprus]. *Tuber Toraks*. 2010; 58(4): 393-400.
- Asut Ö. Smoking status among medical students of Near East University in Nicosia. *Cyprus J Med Sci*. 2019; 4(1): 1-21.
- Aslan D, Ay P, Raymond K, Aşut Ö, Abuduxike G, Şengelen M. Medical students' tobacco consumption status and experiences with smoke-free law violations in enclosed spaces in Türkiye and Northern Cyprus. *Turk Thorac J*. 2025; 26(2): 48-54.
- Asut O, Vaizoglu S, Abuduxike G, Khader E, Ramadan NG, Cali S. Consumption of tobacco products and associated factors among outpatient visitors of two healthcare facilities in Northern Cyprus: a descriptive cross-sectional study. *J Addict Dis*. 2024; 42(4): 1-8.
- Tütün Ürünlerinin Zararlarından Korunma ve Denetim Yasası [Law on Protection from and Control of the Harms of Tobacco Products]. Resmi Gazete [Official Gazette]. 2008; 182.
- World Health Organization. WHO Framework Convention on Tobacco Control: guidelines for implementation. Article 5.3; Article 8; Articles 9 and 10; Article 11; Article 12; Article 13; Article 14. 2013 ed. Geneva: World Health Organization; 2013. Available from: https://www.afro.who.int/sites/default/files/2017-06/9789241505185_eng.pdf
- WHO Regional Office for the Eastern Mediterranean. MPOWER in the Eastern Mediterranean Region. Cairo: WHO EMRO; 2015. Available from: <https://iris.who.int/items/935919fd-fadd-46b5-b9e2-253e1a6f1c8e>
- Abu-Rmeileh NME, Khader YS, Abdul Rahim H, Mostafa A, Nakkash RT, Hamadeh RR, et al. Tobacco control in the Eastern Mediterranean region: implementation progress and persisting challenges. *Tob Control*. 2022; 31(2): 150-2.
- Akter N, Banu B, Chowdhury SH, Tabassum TT, Hossain SM. Compliance status of tobacco control laws in the university setting of Bangladesh: an analytical study followed a mixed-method approach. *BMJ Public Health*. 2024; 2(1): e000496.
- Rijhwani K, Mohanty VR, Balappanavar AY, Hashmi S. Compliance assessment of cigarette and other tobacco products act in public places in Delhi government Hospitals. *Asian Pac J Cancer Prev*. 2018; 19(8): 2097-102.
- ChangeLab Solutions. Model comprehensive smokefree places ordinance [Internet]. Oakland (CA): ChangeLab Solutions; [cited 2025 Aug 24]. Available from: <https://changelabsolutions.org/publications/comp-smokefree-places>

Associations Between Conventional Breast MRI Signal Characteristics and Immunohistochemical Biomarkers in Breast Cancer: A Retrospective Exploratory Study

Uğurcan Balyemez¹, Sinan Akay², Hatice Tuba Sanal³, Mustafa Taşar³

¹Department of Radiology, Near East University Hospital, Nicosia, Cyprus

²Department of Radiology, University of Iowa Health Care, Iowa, United States

³Clinic of Radiology, University of Health Sciences Türkiye, Gülhane Training and Research Hospital, İstanbul, Türkiye

Abstract

BACKGROUND/AIMS: This study investigates whether quantitative magnetic resonance imaging (MRI)-derived signal measurements are associated with immunohistochemical biomarkers in breast cancer, extending beyond morphological assessment to signal-based imaging characteristics.

MATERIALS AND METHODS: The study retrospectively included 63 patients with 66 breast cancer lesions, selected from an initial cohort of 415 patients based on specific inclusion criteria. MRI scans were performed using a standardized protocol, and quantitative MRI signal measurements were extracted and analysed. Histopathological data-including estrogen receptor (ER), progesterone receptor, human epidermal growth factor receptor 2 status, and Ki-67 proliferation index-were also evaluated. Statistical analyses, including correlation and regression tests, were conducted to assess associations between MRI features and immunohistochemical biomarkers.

RESULTS: Weak associations were observed between pre-contrast T1 signal intensity (SI) and ER expression, and between T2 SI and the Ki-67 proliferation index (Spearman's $\rho=-0.17$ and 0.25 , respectively); neither remained statistically significant after Bonferroni correction for multiple comparisons nor after adjustment for within-patient lesion clustering. A more robust positive association was observed between the initial enhancement ratio (IER) and Ki-67 ($\rho=0.39$, 95% confidence interval: $0.16-0.57$, $p=0.001$), which remained statistically significant after Bonferroni correction (corrected $p=0.016$) and in both a single-lesion-per-patient sensitivity analysis and a cluster-robust regression.

CONCLUSION: The association between IER and Ki-67 remained statistically significant after correction for multiple comparisons and after accounting for within-patient lesion clustering, suggesting that it may represent a comparatively robust exploratory imaging correlate of tumor proliferation. Associations involving T1 SI and T2 SI were weak and did not survive correction. These findings are preliminary and hypothesis-generating, underscoring the need for further studies with larger datasets to clarify the clinical relevance of quantitative MRI analysis in breast cancer evaluation.

Keywords: Magnetic resonance imaging, breast cancer, signal intensity, initial enhancement ratio, Ki-67 index

To cite this article: Balyemez U, Akay S, Sanal HT, Taşar M. Associations between conventional breast MRI signal characteristics and immunohistochemical biomarkers in breast cancer: a retrospective exploratory study. Cyprus J Med Sci. 2026;11(5):409-416

ORCID IDs of the authors: U.B. 0000-0001-6845-0081; S.A. 0000-0001-7201-475X; H.T.S. 0000-0003-1899-2813; M.T.0000-0002-4677-3928.



Corresponding author: Uğurcan Balyemez

E-mail: ugurcan.balyemez@neu.edu.tr

ORCID ID: orcid.org/0000-0001-6845-0081

Received: 17.05.2026

Accepted: 04.09.2026

Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.

This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

INTRODUCTION

Breast cancer is the most prevalent malignancy among women and the second leading cause of cancer-related death globally.^{1,2} These striking statistics have made breast cancer one of the most extensively studied topics in modern medicine. This increased attention has led to widespread screening programs, enabling earlier detection, improving treatment protocols, and enhancing survival outcomes.

Understanding the molecular subtypes of breast cancer plays a critical role in facilitating personalized treatment strategies.^{3,4} Gene expression studies have identified several distinct breast cancer subtypes that differ markedly in prognosis and in the therapeutic targets they express; however, these are not applied in daily clinical practice.⁵ Medical oncologists initiate treatment based on the predicted which are largely focused on estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and Ki-67 proliferation levels. There are significant discordances between molecular subtypes and immunohistochemical studies.⁶ However, assessment is costly, necessitating highly specialized equipment and significant technical expertise.

Medical imaging plays a pivotal role in breast cancer screening, diagnosis, treatment planning, and follow-up. Beyond the binary classification of lesions as benign or malignant, imaging parameters obtained through breast magnetic resonance imaging (MRI) may also yield insights into the histopathological subtypes of tumors.

Numerous studies have examined the association between molecular subtypes and morphological imaging features such as lesion shape, margins, spiculation, and location.⁷⁻¹² However, the number of studies examining the relationship between signal properties of MRI and immunohistochemical biomarkers of breast cancer lesions remains limited. In this study, we aim to assess the potential relationship between immunohistochemical markers and their corresponding quantitative MRI-derived parameters.

MATERIALS AND METHODS

Patient Population

This study was approved in University of Health Sciences Non-Interventional Ethics Committee (approval no: 18/84, date: 05.04.2018). All procedures involving human participants were conducted in accordance with the ethical standards of the institutional research committee and with the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. Written informed consent was obtained from all participants for the storage of medical data in the hospital database and its use for research purposes. A retrospective search of the local hospital's medical imaging archive was conducted to identify breast MRI scans performed between January 2014 and April 2018. Imaging data from 415 patients were initially identified. Of these, 91 patients had a histopathological diagnosis of breast cancer. Patients were excluded if they had poor image quality, had a history of neoadjuvant therapy, or lacked detailed histopathological reports. Additionally, cases of *in situ* carcinoma were excluded because these lesions were unsuitable for optimal region of interest (ROI) measurement and immunohistochemical biomarkers were unavailable in their pathology reports. In patients with multiple lesions, all histopathologically confirmed malignant lesions were included in the analysis. After applying exclusion criteria, 63 patients

with a total of 66 breast cancer lesions were included in the final analysis. A study flow diagram, prepared according to STROBE reporting guidelines for observational studies, summarizes the patient and lesion selection process (Figure 1). Exclusion criteria at the n=91 to n=63 stage (poor image quality, neoadjuvant treatment history, incomplete histopathology documentation) were applied concurrently in a single retrospective eligibility screen rather than sequentially. Some excluded patients met more than one criterion; per-criterion counts were not retained and are therefore not reported. Only the total number of exclusions at this stage (n=28) is available.

Imaging Protocol

All breast MRI scans were performed using a standardized protocol, with patients in the prone position, utilizing a 7-channel bilateral breast coil on a 3.0 Tesla superconducting MRI system (Achieva 3.0T X-series, Philips Medical Systems B.V., Eindhoven, The Netherlands). The pre-contrast sequences included axial Turbo Spin Echo T1-weighted (T1w), short tau inversion recovery (STIR), and T2-weighted (T2w) images. Detailed imaging parameters are provided in Table 1. A gadolinium-based contrast agent was administered at a dose of 0.1 mmol/kg, and dynamic contrast-enhanced images were acquired at one-minute intervals for a total of 6 minutes post-injection. MRI data were processed on a dedicated workstation using DynaCad version 2.1.6.81957. All MRI measurements were independently performed by two radiologists with 4 and 10 years' experience in breast imaging, respectively. Interobserver reliability, assessed using the intraclass correlation coefficient (ICC). Using a two-way random-effects, absolute-agreement, single-measures [ICC (2,1)=0.88], indicating good agreement between the two readers; a 95% confidence interval (CI) was not computed for this coefficient. Both radiologists were blinded to each other's results and to the histopathological findings at the time of measurement. Given this

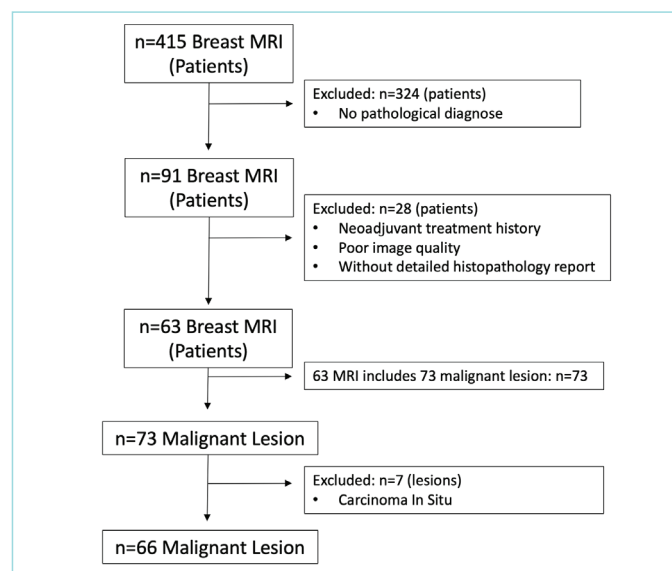


Figure 1. Study flow diagram (prepared according to STROBE reporting guidelines for observational studies). Because eligibility was assessed in a single concurrent screening pass rather than sequentially, per-criterion exclusion counts within this step were not retained and are not reported; only the total number excluded (n=28) is shown.

MRI: Magnetic resonance imaging, STROBE: Strengthening the Reporting of Observational Studies in Epidemiology.

confirmed agreement, the mean of the two radiologists' independent measurements for each lesion was used as the input for subsequent statistical analyses.

Breast MR images and corresponding histopathology reports were reviewed concurrently. Lesions confirmed as malignant by histopathology were identified on MRI, and mean signal intensities were calculated on each sequence using elliptical ROIs manually placed on the single axial slice that demonstrated the largest cross-sectional area of each lesion, encompassing at least two-thirds of that area. Areas of necrosis or hemorrhage were excluded from the ROIs to avoid signal intensity (SI) alterations unrelated to tumor tissue (Figure 2). Signal intensities were measured on unenhanced T1w and T2w sequences and compared with histopathological parameters. Additionally, the initial enhancement ratio (IER) was calculated as the percentage increase in T1 SI between the pre-contrast image and the first post-contrast time point, using the same ROI (Figure 3) to minimize measurement bias. $IER (\%) = [(SI \text{ post-contrast, first minute} - SI \text{ pre-contrast}) / SI \text{ pre-contrast}] \times 100$.

Histopathologic Assessment

Histopathological data were retrieved from patients' medical records and evaluated retrospectively. As this was a retrospective study, histopathological evaluations were performed by the pathologists on duty at the time of each biopsy, in accordance with the institution's routine clinical workflow. All reports were generated using standardized immunohistochemical staining protocols applied consistently across the study period. The following features were assessed: tumor histological subtype, grade, ER, PR, HER2 status, and Ki-67 proliferation index. Molecular subtypes were categorized as Luminal A, Luminal B, HER2-enriched, and Basal-like. Molecular classification was based on a combination of tumor grade, ER, PR, HER2 status, and Ki-67 index. HER2 status was extracted from each patient's final composite pathology report rather than from the immunohistochemical score alone; immunohistochemically equivocal (2+) cases were reflex-tested with fluorescence *in situ* hybridization (FISH) as part of routine clinical practice during the study period, consistent with ASCO/CAP guidelines, and the FISH result was incorporated into the final reported HER2 status. HER2 was therefore analyzed as a binary variable (positive vs. negative).

Table 1. Detailed imaging parameters of the MRI sequences			
	Pre-contrast T1	T2 STIR	Dynamic 3D T1
Slice thickness	3 mm	3 mm	0.76/0.76/2 mm
Gap	0 mm	0 mm	0 mm
Matrix	340×277	340×259	472×472
FOV (AP × RL × FH)	340×352×150	340×340×150	360×360×150
TR	495 ms	10927 ms (TI=230 ms)	5.1 ms
TE	10 ms	60 ms	2.6 ms
NEX	1	1	1
Time (min:sec)	3.21	4.22	1.07
Total time (min:sec)	-	-	7.52

MRI: Magnetic resonance imaging, STIR: Short tau inversion recovery, FOV: Field of view, AP: Anteroposterior, RL: Right-left, FH: Foot-head, TR: Repetition time, TE: Echo time, TI: Inversion time, NEX: Number of excitations.

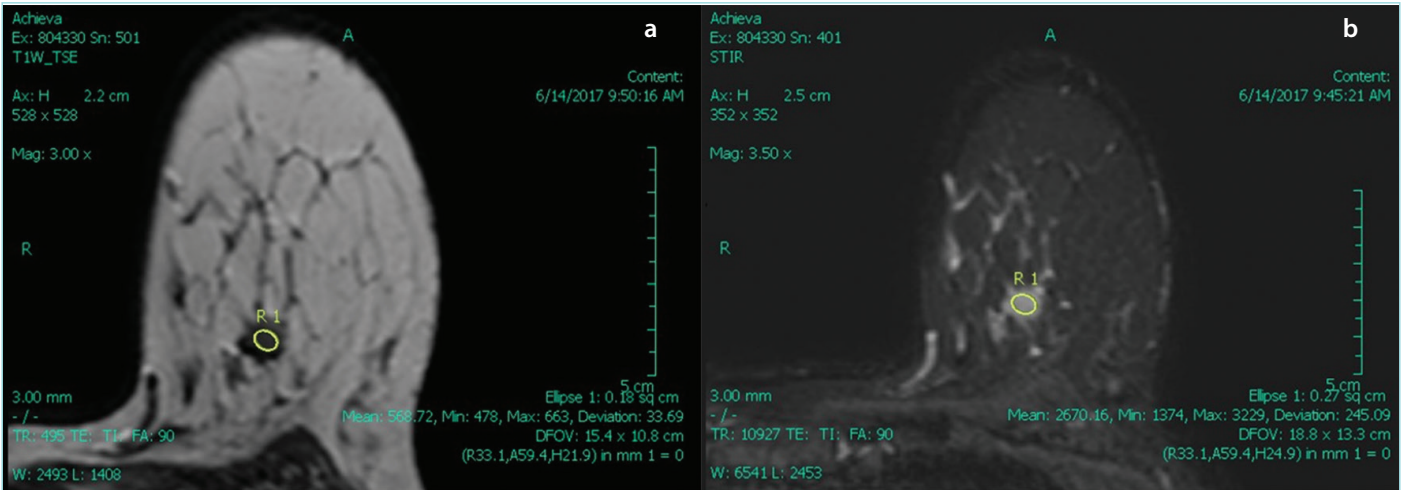


Figure 2. Mean signal calculations by using an ROI on T1w axial (a) and T2w axial (b) images (mean T1 SI on a=568.72, and mean T2 SI on b=2670.16).
ROI: Region of interest, T1w: T1-weighted, T2w: T2-weighted, SI: Signal intensity.

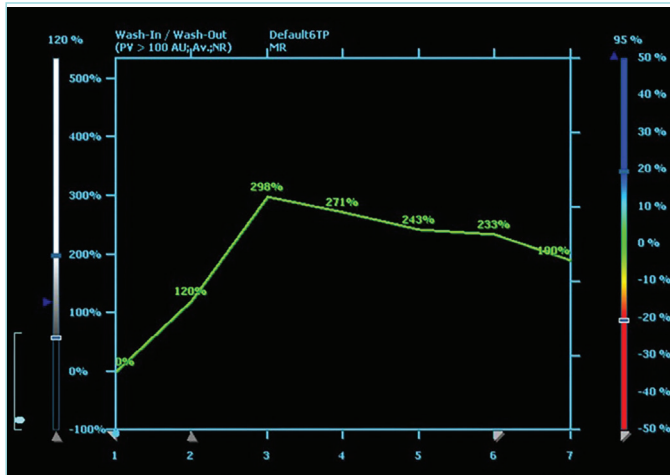


Figure 3. Calculation of initial enhancement ratio by using the time-intensity curve at the dedicated workstation. Time point 1 corresponds to the unenhanced T1 signal intensity, and time point 2 shows the first-minute signal change after contrast material injection. In this example, IER was 120%.

IER: Initial enhancement ratio.

Molecular subtypes were assigned using standard literature-based criteria: ER and PR positivity were defined as $\geq 1\%$ positively stained tumor nuclei by immunohistochemistry, consistent with ASCO/CAP reporting standards; a Ki-67 cut-off of 20% was used to distinguish low- from high-proliferation tumors, consistent with the St Gallen International Expert Consensus framework; HER2 positivity was determined according to ASCO/CAP guideline criteria, as described above; and histological grade was incorporated as a supportive criterion, consistent with the St Gallen consensus framework. These thresholds were applied solely for the descriptive molecular subtype classification; the primary correlation analyses (Table 2) used the continuous values of ER, PR, and Ki-67 directly, without dichotomization. Given the resulting skewed subtype distribution in this cohort-including a single basal-like lesion and no HER2-enriched lesions-comparisons across molecular subtypes were not adequately powered for formal statistical testing and

were presented as descriptive and exploratory findings only, rather than as confirmatory subgroup analyses.

Statistical Analysis

The Shapiro-Wilk test was applied to evaluate the normality of the distributions of all imaging (T1 SI, T2 SI, IER) and biomarker (ER, PR, Ki-67) variables. Only patient age was normally distributed ($p=0.559$); all other variables significantly deviated from normality (all $p<0.05$). Spearman's rank correlation coefficient was therefore used as the primary measure of association for all twelve pre-specified MRI parameter-biomarker combinations (T1 SI, T2 SI, and IER, each against ER, PR, HER2, and Ki-67) continuous; this approach treated HER2 as a two-level ordinal variable for this purpose, consistent with the ordinal/non-parametric treatment applied to the other markers. Because some patients contributed more than one lesion to the dataset, within-patient dependence was addressed using (1) a sensitivity analysis restricted to a single index lesion per patient, and (2) cluster-robust regression with standard errors (SEs) clustered at the patient level. Multiple linear regression was used to assess the combined association of T2 SI and IER with Ki-67, including collinearity diagnostics (variance inflation factor) and assessment of residuals. All statistical analyses were performed using STATA version 14 (StataCorp, College Station, TX, USA), with supplementary Spearman and cluster-robust analyses conducted in Python (SciPy/statsmodels). A p -value <0.05 was considered statistically significant. To account for the twelve bivariate correlation tests performed, a Bonferroni-corrected significance threshold of $\alpha=0.0042$ ($0.05/12$) was applied.

As the multiple regression residuals showed some deviation from normality, Spearman's rank correlation was retained as the primary analysis throughout, with linear regression presented as a secondary, exploratory analysis. A bootstrap resampling analysis (10,000 resamples, patient-level clustering) was additionally performed on the primary IER-Ki-67 correlation as a non-parametric robustness check.

RESULTS

A total of 63 patients with 66 malignant breast lesions were included in a retrospective analysis. The mean age of the patients was 49.5 ± 9 years

Table 2. Summary of bivariate correlations between MRI signal parameters and immunohistochemical biomarkers

MRI parameter	IHC marker	n	Spearman ρ	95% CI	p (uncorrected)	p (Bonferroni $\times 12$)
T1 SI	ER	66	-0.168	-0.39 to 0.08	0.177	1.000
T1 SI	PR	66	-0.138	-0.37 to 0.11	0.270	1.000
T1 SI	HER2	66	0.078	-0.17 to 0.31	0.535	1.000
T1 SI	Ki-67	66	0.124	-0.12 to 0.36	0.321	1.000
T2 SI	ER	66	-0.063	-0.30 to 0.18	0.618	1.000
T2 SI	PR	66	0.097	-0.15 to 0.33	0.436	1.000
T2 SI	HER2	66	-0.012	-0.25 to 0.23	0.924	1.000
T2 SI	Ki-67	66	0.250	0.01 to 0.46	0.043	0.513
IER	ER	66	-0.182	-0.41 to 0.06	0.143	1.000
IER	PR	66	-0.045	-0.28 to 0.20	0.719	1.000
IER	HER2	66	0.132	-0.11 to 0.36	0.290	1.000
IER	Ki-67	66	0.386	0.16 to 0.57	0.001	0.016

Spearman's rank correlation was used given non-normal distributions (Shapiro-Wilk testing). All twelve pre-specified MRI parameter-biomarker combinations were tested; the Bonferroni-corrected significance threshold was $\alpha=0.0042$ ($0.05/12$). Only the IER-Ki-67 association retained significance after correction.

MRI: Magnetic resonance imaging, SI: Signal intensity, IHC: Immunohistochemistry, CI: Confidence interval, IER: Initial enhancement ratio, ER: Estrogen receptor, PR: Progesterone receptor, HER2: Human epidermal growth factor receptor 2.

(range: 31-71 years). The majority of the lesions (68.4%) were classified as invasive ductal carcinoma. Based on immunohistochemical analysis, the molecular subtypes of the lesions were distributed as follows: Luminal A (n=18, 27.3%); Luminal B (n=47, 71.2%); and Basal-like (n=1, 1.5%). No lesions were categorized as HER2-enriched. Statistical analysis revealed no significant associations between quantitative MRI signal measurements and the molecular subtypes ($p>0.05$).

Pre-contrast T1 SI values ranged from 250 to 824, with a median of 375 SI and a mean of 391.8 ± 111.5 SI. T2w STIR SI values ranged from 303 to 3273 SI, with a median of 975 SI and a mean of 1035.2 ± 490.5 SI. IER values ranged from 10% to 220%, with a median of 91% and a mean of $101.4 \pm 60.7\%$.

Among the histopathological findings, Ki-67 values ranged from 2% to 80%, with a median value of 15% and a mean of 19.7 ± 15.2 . ER expression varied between 0% and 100%, with a median of 90% and a mean of 74.2 ± 31.3 . PR values also ranged from 0% to 100%, with a median of 40% and a mean of 44 ± 35.5 . The associations between MRI parameters and the expression levels of ER, PR, and HER2 were evaluated individually. A weak negative correlation between pre-contrast SI values and ER expression was observed (Spearman's $\rho = -0.17$, 95% CI: -0.39 to 0.08, $p = 0.177$), but it did not reach statistical significance (Figure 4a).

T2w STIR SI values showed no significant correlation with ER, PR, or HER2 expression levels (all Spearman's ρ , uncorrected $p > 0.05$; Table 2). A weak positive correlation was observed between T2 SI values and the Ki-67 proliferation index ($\rho = 0.25$, 95% CI: 0.01 to 0.46, $p = 0.043$), but it did not remain significant after Bonferroni correction for multiple comparisons (corrected $p = 0.513$) nor in cluster-robust regression ($p = 0.105$) (Figure 4b).

Analysis of IER values revealed no significant associations between IER and ER, PR, or HER2 status (all Spearman's ρ , uncorrected $p > 0.05$; Table 2). In contrast, a moderate positive correlation was found between IER and Ki-67 levels ($\rho = 0.39$, 95% CI: 0.16 to 0.57, $p = 0.001$), which remained statistically significant after Bonferroni correction for multiple comparisons (corrected $p = 0.016$), in a single-lesion-per-patient sensitivity analysis ($\rho = 0.43$, $p < 0.001$), and in cluster-robust regression ($p = 0.028$) (Figure 4c).

Multiple linear regression of Ki-67 on T2 SI and IER (n=66) showed that IER was independently associated with Ki-67 in the model ($B = 0.077$, $SE = 0.030$, 95% CI: 0.017-0.136, $p = 0.012$), whereas T2 SI was not independently associated ($B = 0.006$, $SE = 0.004$, 95% CI: -0.001-0.013, $p = 0.117$). The overall model was statistically significant [$F(2,63) = 5.54$, $p = 0.006$; $R^2 = 0.150$, adjusted $R^2 = 0.123$], with minimal collinearity between predictors (variance inflation factor = 1.04 for both). When SEs were adjusted for within-patient clustering, IER remained significant (cluster-robust $p = 0.028$), whereas T2 SI did not ($p = 0.178$). Residuals showed a deviation from normality (Shapiro-Wilk $p < 0.001$), consistent with the underlying skewed distribution of Ki-67. Results were nonetheless consistent across parametric and non-parametric approaches.

As an additional non-parametric sensitivity check, a bootstrap resampling analysis (10,000 resamples) was performed to evaluate the Spearman correlation between IER and Ki-67. A patient-level cluster bootstrap, resampling patients with replacement to account for within-patient lesion dependence, yielded a 95% CI of 0.12 to 0.62; a simple case-resampling bootstrap yielded a 95% CI of 0.13 to 0.60. Both intervals excluded zero and were consistent with the Fisher z-transformation-based CI reported above, further supporting the robustness of this association.

A complete summary of twelve bivariate correlations between MRI signal parameters and immunohistochemical biomarkers, including exact coefficients, 95% CIs, and both uncorrected and Bonferroni-corrected p-values, is presented in Table 2. After Bonferroni correction for multiple comparisons ($\alpha = 0.0042$), only the association between IER and Ki-67 retained statistical significance; the T1 SI-ER and T2 SI-Ki-67 associations did not retain statistical significance and should be interpreted as exploratory.

DISCUSSION

Recent studies in breast radiology, as well as globally accepted guidelines (e.g., BI-RADS), aim to assess the malignancy risk of detected lesions.¹³ Advances in imaging technology and computational analysis have allowed us to establish diverse associations with the pathological features of tumors beyond the differentiation of benign or malignant lesions.

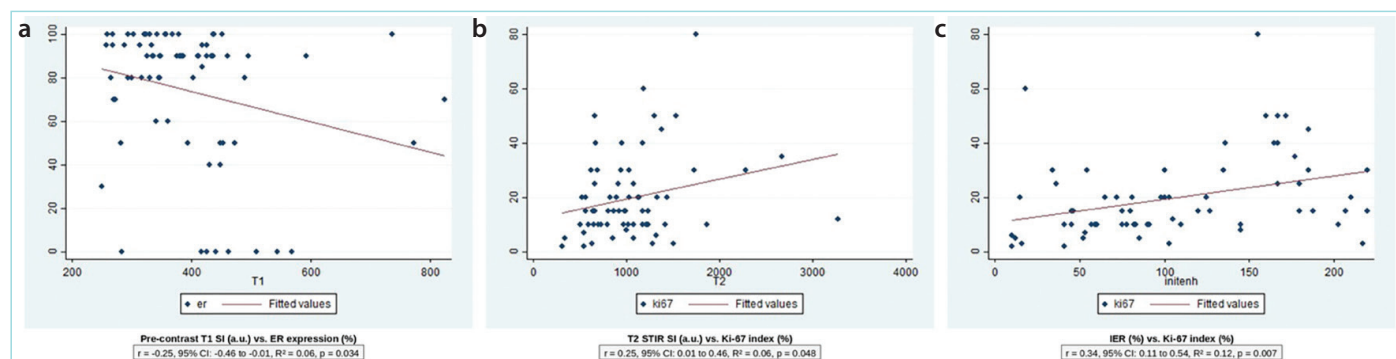


Figure 4. Bivariate correlations between MRI signal parameters and immunohistochemical biomarkers (Spearman's rank correlation). (a) Correlation between pre-contrast T1 SI values and ER expression (n=66, $\rho = -0.17$, 95% CI: -0.39 to 0.08, $p = 0.177$). (b) Correlation between T2 SI values and Ki-67 index (n=66, $\rho = 0.25$, 95% CI: 0.01 to 0.46, $p = 0.043$). (c) Correlation between initial enhancement ratio and Ki-67 index (n=66, $\rho = 0.39$, 95% CI: 0.16 to 0.57, $p = 0.001$).

IER: Initial enhancement ratio, ER: Estrogen receptor, MRI: Magnetic resonance imaging, SI: Signal intensity, STIR: Short tau inversion recovery, CI: Confidence interval.

Several recent studies have explored the role of MRI features in identifying breast cancer molecular subtypes.^{10-12,14-17} For instance, multiparametric MRI has demonstrated significant differences among subtypes, particularly between triple-negative and hormone receptor-positive tumors.¹⁸ Moreover, dynamic contrast-enhanced MRI radiomics, especially intratumoral and peritumoral features derived from late-phase images, have been shown to contribute substantially to subtype classification,¹⁹ while multiparametric approaches overall outperform semantic models in predicting molecular subtypes.²⁰ In addition, morphological characteristics such as mass shape and margin features have also been reported to correlate with specific subtypes, including round contours in triple-negative tumors and spiculated margins in hormone receptor-positive cancers.²¹ However, in our study, we did not observe significant associations of enhancement patterns or SI values (T1 SI, T2 SI, or IER) with molecular subtypes. This discrepancy may be related to differences in imaging protocols, sample size, or in the use of conventional imaging parameters rather than advanced radiomic or multiparametric approaches. Our findings suggest that while MRI has potential in molecular subtype assessment, classical enhancement and SI measurements alone may be insufficient for reliable subtype differentiation in routine clinical practice. This lack of association should also be interpreted in light of the skewed molecular subtype distribution in our cohort, which may limit the generalizability of this particular finding and is discussed further in the study limitations section.

We also examined the association between pre-contrast T1 SI values and ER expression (Spearman's $\rho = -0.17$, $p = 0.177$), which did not reach statistical significance after correction for multiple comparisons (see Table 2). While post-contrast T1 signal variability has been described in the literature, the relevance of pre-contrast T1 SI to tumor biology remains underexplored.²² To our knowledge, no previous study has specifically investigated the T1 SI of fibroglandular tissue or breast lesions on pre-contrast T_w images. In our cohort, lower T1 SI values were associated with higher ER expression. Previous studies have demonstrated lower ADC values in ER-positive tumors, reflecting higher cellularity.^{23,24} Given that this association did not remain statistically significant after Bonferroni correction for multiple comparisons or adjustment for within-patient lesion clustering (Spearman's $\rho = -0.17$, $p = 0.177$; see statistical analysis and Table 2), the association should be interpreted as a non-significant, exploratory observation rather than evidence of biomarker capability and the above hypothesis-generating interpretation should therefore be considered preliminary.

In the current study, we also examined the T2 STIR signal intensities of malignant lesions. While early studies such as Kuhl et al.²⁵ reported no statistically significant differentiation between benign and malignant lesions based on T2 STIR imaging, more recent evidence offers a different perspective. Notably, Kim et al.²⁶ found that intratumoral T2 hyperintensity correlated positively with Ki-67 in ER-positive/HER2-negative invasive ductal carcinoma, while ADC values did not show significant associations. Dietzel et al.²⁷ emphasized that T2w hyperintense signal is associated with higher proliferation rates and histologic grade, suggesting its potential as a non-contrast imaging biomarker. In our study, we similarly found that higher lesion SI values on T2 STIR were positively associated with the Ki-67 proliferation index. This observation is in line with the findings of Kim et al.²⁶ who reported that intratumoral T2 hyperintensity correlated with high Ki-67 expression in ER-positive/HER2-negative breast cancers, and with the

work of Dietzel et al.,²⁷ who demonstrated that T2 hyperintensity and other semantic features were linked to increased proliferative activity and higher histologic grade. Taken together, these results should be interpreted cautiously, as the association did not survive correction for multiple comparisons or patient-level clustering (Table 2).

Recent studies have highlighted the role of dynamic contrast-enhanced MRI kinetics in reflecting tumor proliferative activity. Szep et al.¹⁸ demonstrated that multiparametric MRI features, including kinetic parameters, show significant associations with molecular subtypes and proliferation indices. Similarly, Huang et al.¹⁹ reported that radiomic features derived from late-phase DCE-MRI, particularly intratumoral and peritumoral characteristics, contributed substantially to Ki-67 prediction, while Huang et al.²⁰ confirmed the superiority of multiparametric MRI over semantic models in this context. In our study, IER showed a statistically significant association with Ki-67 (Spearman's $\rho = 0.39$, $p = 0.001$). This association remained significant after Bonferroni correction for multiple comparisons (corrected $p = 0.016$), in a single-lesion-per-patient sensitivity analysis, and in cluster-robust regression accounting for within-patient lesion dependence. Taken together, these findings suggest that enhancement kinetics may serve as a non-invasive imaging biomarker for tumor proliferation, providing complementary prognostic information. However, MRI-derived parameters, including IER, cannot currently replace tissue-based immunohistochemical assessment and should be regarded as complementary exploratory signals that require prospective validation.

Weak, uncorrected relationships were observed between T1 SI and ER, and between T2 SI and Ki-67; however, neither remained significant after correction for multiple comparisons or adjustment for within-patient lesion clustering. Both should be interpreted as exploratory. In contrast, the relationship between IER and Ki-67 remained statistically significant in Spearman correlation with Bonferroni correction, in a single-lesion-per-patient sensitivity analysis, and in cluster-robust regression, representing a comparatively robust finding. An additional strength of our study was that IER could be calculated using dedicated software with minimal operator dependence, providing a quantitative and reproducible parameter. The objective nature of IER may support its role as an exploratory imaging correlate, though it cannot currently replace tissue-based immunohistochemical assessment. Overall, the MRI parameters evaluated in this study represent only a small component of a broader research field. However, our findings provide preliminary insights into the histopathological characteristics of malignant breast lesions beyond the classical benign-malignant differentiation and warrant further investigation.

Study Limitations

First, although the signal measurements in our study were performed at the dedicated workstation, they might not reflect the whole lesion due to operator-dependent ROI placement, which may not encompass the entire lesion. Second, all patients with malignant histopathology results were evaluated as a single group. Although the vast majority (68.49%), breast cancer subtypes and histopathological biomarkers may affect the MRI findings we described. Furthermore, the molecular subtype distribution in our cohort was heavily skewed toward Luminal A and Luminal B tumors, with only a single basal-like lesion and no HER2-enriched lesions; findings regarding molecular subtype should therefore not be generalized to underrepresented or absent subtypes. Third, the relatively small sample size (63 patients and 66 lesions)

limits the statistical power and generalizability of our findings. Future multicenter studies with larger cohorts are warranted. Additionally, because histopathological assessments were performed by multiple pathologists during routine clinical care, interobserver variability could not be controlled. A further limitation is the use of absolute, non-normalized SI values rather than SI ratios referenced to a stable tissue (e.g., pectoralis muscle or contralateral fibroglandular tissue). Absolute SI is influenced by scanner hardware, receiver gain, coil sensitivity, and acquisition settings, which may limit direct comparability across examinations. Although all scans were acquired on a single scanner using an identical, standardized protocol, prospective studies should incorporate reference-tissue normalization to improve reproducibility. Finally, although Spearman's rank correlation was used as the primary method following Shapiro-Wilk normality testing (see statistical analysis), residual non-normality in the multiple regression model (Shapiro-Wilk $p < 0.001$ for residuals) indicates that parametric regression estimates should be interpreted with some caution. However, results were consistent with the non-parametric Spearman correlation and cluster-robust analyses. Diagnostic threshold determination (e.g., via ROC analysis to identify a clinically actionable Ki-67 cut-off) requires a study specifically designed and powered for this purpose and is beyond the scope of the present exploratory analysis. This represents an important direction for future prospective, adequately powered studies.

CONCLUSION

In summary, the association between the IER and Ki-67 remains statistically significant after Bonferroni correction for multiple comparisons and after accounting for within-patient lesion clustering, and represents the most robust finding of this study. Associations between pre-contrast T1 SI and ER, and between T2 SI and Ki-67, were weak and did not survive correction. These findings should be regarded as preliminary and hypothesis-generating rather than as established imaging biomarkers, using normalized SI measures before any consideration of clinical implementation. Moreover, the increasing interest in digitalization, artificial intelligence, and quantitative data assessment indicates that radiomic features may play an important role in breast cancer evaluation in the near future.

MAIN POINTS

- A weak positive correlation was observed between T2 short tau inversion recovery signal and tumor Ki-67 expression levels, which did not remain significant after correction for multiple comparisons or for patient-level clustering.
- A correlation between initial enhancement ratio (IER) and tumor Ki-67 values was identified and remained statistically significant after Bonferroni correction and in sensitivity analyses accounting for within-patient lesion clustering. This finding suggests that IER may be a comparatively robust exploratory imaging correlate of Ki-67 expression, which provides important information about tumor prognosis and mitotic activity.
- Future studies involving larger datasets and analyzing additional magnetic resonance parameters could improve understanding of associations with immunohistochemical markers.

ETHICS

Ethics Committee Approval: This study was approved in University of Health Sciences Non-Interventional Ethics Committee (approval no: 18/84, date: 05.04.2018).

Informed Consent: Written informed consent was obtained from all participants for the storage of medical data in the hospital database and its use for research purposes.

Footnotes

Authorship Contributions

Concept: M.T., Design: H.T.S., Data Collection and/or Processing: U.B., Analysis and/or Interpretation: U.B., S.A., M.T., Literature Search: U.B., S.A., H.T.S., Writing: U.B., H.T.S.

DISCLOSURES

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

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

Declaration Regarding the Use of AI and AI-Assisted Technologies:

During the preparation of this manuscript, the authors used Claude (Anthropic) and ChatGPT (OpenAI) for sentence-level editing and grammatical refinement. All AI-generated suggestions were carefully reviewed, critically evaluated, and edited as necessary by the authors to ensure accuracy and contextual appropriateness. The authors take full responsibility for the publication's content. The use of AI tools primarily affected the linguistic presentation and grammatical structure of the manuscript; all scientific content, data interpretation, and conclusions reflect solely the authors' own work.

REFERENCES

1. Giaquinto AN, Sung H, Newman LA, Freedman RA, Smith RA, Star J, et al. Breast cancer statistics 2024. *CA Cancer J Clin.* 2024; 74(6): 477-95.
2. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin.* 2024; 74(1): 12-49.
3. Dai X, Li T, Bai Z, Yang Y, Liu X, Zhan J, et al. Breast cancer intrinsic subtype classification, clinical use and future trends. *Am J Cancer Res.* 2015; 5: 2929-43.
4. Gradishar WJ, Moran MS, Abraham J, Abramson V, Aft R, Agnese D, et al. Breast cancer, version 3.2024. NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw.* 2024; 22: 331-57.
5. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature.* 2012; 490(7418): 61-70.
6. Kim HK, Park KH, Kim Y, Park SE, Lee HS, Lim SW, et al. Discordance of the PAM50 intrinsic subtypes compared with immunohistochemistry-based surrogate in breast cancer patients: potential implication of genomic alterations of discordance. *Cancer Res Treat.* 2019; 51(2): 737-47.
7. Li H, Zhu Y, Burnside ES, Huang E, Drukker K, Hoadley KA, Fan C, et al. Quantitative MRI radiomics in the prediction of molecular classifications of breast cancer subtypes in the TCGA/TICIA data set. *NPJ Breast Cancer.* 2016; 2: 16012-.
8. Bae MS, Seo M, Kim KG, Park IA, Moon WK. Quantitative MRI morphology of invasive breast cancer: correlation with immunohistochemical biomarkers and subtypes. *Acta Radiol.* 2015; 56(3): 269-75.

9. Whitney HM, Taylor NS, Drukker K, Edwards AV, Papaioannou J, Schacht D, et al. Additive benefit of radiomics over size alone in the distinction between benign lesions and luminal a cancers on a large clinical breast MRI dataset. *Acad Radiol*. 2019; 26(2): 202-9.
10. Ab Mumin N, Ramli Hamid MT, Wong JHD, Rahmat K, Ng KH. Magnetic resonance imaging phenotypes of breast cancer molecular subtypes: a systematic review. *Acad Radiol*. 2022; 29(Suppl 1): S89-106.
11. Cho N. Imaging features of breast cancer molecular subtypes: State of the art. *J Pathol Transl Med*. 2020; 55: 16.
12. Fan M, Li H, Wang S, Zheng B, Zhang J, Li L. Radiomic analysis reveals DCE-MRI features for prediction of molecular subtypes of breast cancer. *PLoS One*. 2017; 12: e0171683.
13. D'Orsi CJ, Sickles EA, Mendelson EB, Morris EA. ACR BI-RADS Atlas: breast imaging reporting and data system. Reston, VA: American College of Radiology; 2013.
14. Yamaguchi K, Abe H, Newstead GM, Egashira R, Nakazono T, Imaizumi T, et al. Intratumoral heterogeneity of the distribution of kinetic parameters in breast cancer: comparison based on the molecular subtypes of invasive breast cancer. *Breast Cancer*. 2015; 22(5): 496-502.
15. Mazurowski MA, Zhang J, Grimm LJ, Yoon SC, Silber JL. Radiogenomic analysis of breast cancer: Luminal B molecular subtype is associated with enhancement dynamics at MR imaging. *Radiology*. 2014; 273(2): 365-72.
16. Grimm LJ, Zhang J, Mazurowski MA. Computational approach to radiogenomics of breast cancer: Luminal A and Luminal B molecular subtypes are associated with imaging features on routine breast MRI extracted using computer vision algorithms. *J Magn Reson Imaging*. 2015; 42(4): 902-7.
17. Blaschke E, Abe H. MRI phenotype of breast cancer: Kinetic assessment for molecular subtypes. *J Magn Reson Imaging*. 2015; 42(4): 920-4.
18. Szep M, Pintican R, Boca B, Perja A, Duma M, Feier D, et al. Multiparametric MRI features of breast cancer molecular subtypes. *Medicina (Kaunas)*. 2022; 58(12): 1716.
19. Huang G, Du S, Gao S, Guo L, Zhao R, Bian X, et al. Molecular subtypes of breast cancer identified by dynamically enhanced MRI radiomics: the delayed phase cannot be ignored. *Insights Imaging*. 2024; 15(1): 127.
20. Huang Y, Wang X, Cao Y, Li M, Li L, Chen H, et al. Multiparametric MRI model to predict molecular subtypes of breast cancer using Shapley additive explanations interpretability analysis. *Diagn Interv Imaging*. 2024; 105(5): 191-205.
21. Temerik SM, Elwahab SMA, Wahman MM, Ahmed MY, Elwanis MEA. Relation between morphological features of initial breast MRI and breast cancer molecular subtypes. *Egypt J Radiol Nucl Med*. 2023; 54: 147.
22. Delille JP, Slanetz PJ, Yeh ED, Kopans DB, Garrido L. Physiologic changes in breast magnetic resonance imaging during the menstrual cycle: perfusion imaging, signal enhancement, and influence of the T1 relaxation time of breast tissue. *Breast J*. 2005; 11(4): 236-41.
23. Kamitani T, Matsuo Y, Yabuuchi H, Fujita N, Nagao M, Jinnouchi M, et al. Correlations between apparent diffusion coefficient values and prognostic factors of breast cancer. *Magn Reson Med Sci*. 2013; 12(3): 193-9.
24. Choi SY, Chang YW, Park HJ, Kim HJ, Hong SS, Seo DY. Correlation of the apparent diffusion coefficient values on diffusion-weighted imaging with prognostic factors for breast cancer. *Br J Radiol*. 2012; 85(1016): e474-9.
25. Kuhl CK, Klaschik S, Mielcarek P, Gieseke J, Wardelmann E, Schild HH. Do T2-weighted pulse sequences help with the differential diagnosis of enhancing lesions in dynamic breast MRI? *J Magn Reson Imaging*. 1999; 9(2): 187-96.
26. Kim SY, Kim EK, Moon HJ, Yoon JH, Koo JS, Kim SG, et al. Association among T2 signal intensity, necrosis, ADC and Ki-67 in estrogen receptor-positive and HER2-negative invasive ductal carcinoma. *Magn Reson Imaging*. 2018; 54: 176-82.
27. Dietzel M, Trimboli RM, Zanardo M, Schultz-Wendtland R, Uder M, Clauser P, et al. The potential of predictive and prognostic breast MRI (P2-bMRI). *Eur Radiol Exp*. 2022; 6(1): 42.

Screen Exposure and Socio-Demographic Characteristics of Children Aged 2-6 Years Diagnosed with Speech and Language Disorder: A Descriptive Study

✉ Bari Ay¹, ✉ Umut Balatacı², ✉ Muzaffer Tunahan Gedikçi³

¹Clinic of Child and Adolescent Psychiatry, Karabük, Türkiye

²Department of Child and Adolescent Psychiatry, Biruni University Faculty of Medicine, İstanbul, Türkiye

³Department of Speech and Language Therapy, Ankara Medipol University Institute of Health Sciences, Ankara, Türkiye

Abstract

BACKGROUND/AIMS: Screen exposure during early childhood may influence cognitive and language development. This descriptive and exploratory study aimed to characterize screen exposure patterns and selected sociodemographic, perinatal, and clinical features of children aged 2-6 years diagnosed with speech and language disorder.

MATERIALS AND METHODS: This descriptive, exploratory, cross-sectional study was conducted between November 2025 and March 2026 at the Child and Adolescent Psychiatry Clinic of Karabük Training and Research Hospital. Children who, following comprehensive psychiatric evaluation, met the diagnostic criteria for speech and language disorder were included. Children with conditions that could better explain their language difficulties, including autism spectrum disorders and intellectual disability, were excluded.

RESULTS: The study included 43 children, of whom 62.8% were male. The mean age was 47.18 ± 15.22 months. Mean gestational age and birth weight were 36.67 ± 2.74 weeks and 3068 ± 727 g, respectively. Cesarean delivery, incubator care, postnatal hypoxia, and prolonged jaundice were reported in 72.1%, 32.6%, 23.3%, and 9.3% of cases, respectively. Daily screen exposure was <1 hour for 37.2%, 1-2 hours for 34.9%, 2-4 hours for 23.3%, and >4 hours for 4.7%. First screen exposure occurred before 12 months in 32.6% of participants, and before 24 months in 74.4%.

CONCLUSION: The findings provide a descriptive overview of screen exposure and selected characteristics in this sample, but do not establish an association with speech and language disorder.

Keywords: Speech and language disorder, screen, communication

INTRODUCTION

The relationship between screen use (exposure to television or other digital screens) and early childhood development has long been discussed in both the scientific literature and the public sphere. Over the past decade, this topic has received renewed attention because of the substantial increase in young children's access to and use of digital media. In the literature, debates have mainly focused on the quantity of screen use (duration of use) and its quality (content and context).

Regarding quantity, prolonged screen exposure has been suggested to function as a passive, sedentary activity that may replace learning opportunities that are critical for child development.¹ Especially in early childhood, the limited reciprocal verbal interaction during screen use is considered a potential risk factor for language acquisition.^{2,3}

Language delay in early childhood is a critical neurodevelopmental concern that may affect not only communication skills but also long-term academic achievement and psychosocial functioning. Current

To cite this article: Ay B, Balatacı U, Gedikçi MT. Screen exposure and socio-demographic characteristics of children aged 2-6 years diagnosed with speech and language disorder: a descriptive study. Cyprus J Med Sci. 2026;11(5):417-423

ORCID IDs of the authors: B.A. 0000-0002-9740-3716; U.B. 0000-0002-1562-1673; M.T.G. 0009-0001-7052-0576.



Corresponding author: Umut Balatacı
E-mail: umutbltcm@gmail.com
ORCID ID: orcid.org/0000-0002-1562-1673

Received: 06.04.2026
Accepted: 11.08.2026
Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.
This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

literature indicates that these effects are more pronounced and persistent when the delay is prolonged.

Longitudinal studies have reported that children with early language delay continue to show significant language-related difficulties compared with their peers during the school-age years. In particular, children with persistent language delay may experience difficulties in mathematics and overall academic performance, in addition to continuing language-related challenges, throughout the primary school years. Moreover, psychosocial problems such as attention difficulties, externalizing behaviors, and peer relationship problems have been reported more frequently in this group.⁴

Since language skills form the foundation of early literacy development, delays in this area directly affect reading, writing, and comprehension skills. Such delays may make it more difficult for children to understand instructions in educational settings and to process academic content, thereby adversely affecting learning. Indeed, children with early language delay have been reported to be at higher risk for later learning difficulties.⁵

From a psychosocial perspective, children with delayed language development may experience limitations in peer interactions, reduced social participation, and behavioral problems due to communication-based frustration.⁶ These challenges may disrupt social adaptation and, in the long term, predispose affected children to social withdrawal, low self-esteem, and emotional difficulties. Furthermore, long-term follow-up studies have shown that individuals with early language difficulties are more likely to experience academic and mental health problems extending into adulthood.⁷

Importantly, the literature emphasizes the value of early intervention. Language and speech interventions initiated before the age of 3 years have been reported to provide significant improvements in receptive and expressive language skills and to reduce long-term adverse outcomes. These findings highlight the critical importance of early diagnosis and a multidisciplinary approach.^{8,9}

Although previous studies have examined screen exposure in relation to language development, the present study provides a structured description of screen use and selected socio-demographic, perinatal, and clinical characteristics in children evaluated in a tertiary child psychiatry setting.¹⁰⁻¹³ It was designed to characterize children diagnosed with speech and language disorder, rather than to establish a causal relationship between screen exposure and language delay. Describing these characteristics within a tertiary clinical setting may help identify factors that should be routinely considered during comprehensive developmental assessment and may generate hypotheses for future controlled studies.

This study aimed to describe screen-use characteristics, including daily screen duration and age at first exposure, and selected socio-demographic and perinatal characteristics among children aged 2-6 years diagnosed with speech and language disorder. Because the study did not include a typically developing control group, it was not designed to determine whether screen exposure differed from that in the general pediatric population or was associated with the development of language delay.

MATERIALS AND METHODS

The study was initiated after obtaining approval from the Karabük University Non-Interventional Clinical Research Ethics Committee (decision no: 2025/2509, date: 30.10.2025). This descriptive, exploratory, cross-sectional study was conducted from November 2025 to March 2026 at the Child and Adolescent Psychiatry Clinic of Karabük Training and Research Hospital. A total of 43 children were included. Written informed consent was obtained from the parents or legal guardians of all participants prior to enrollment. A socio-demographic data form, developed by the researchers to assess socio-demographic characteristics and screen-use patterns, was administered to the participants. Children aged 2-6 years who met the diagnostic criteria for speech and language disorder following psychiatric evaluation were included in the study.

Referral Pathways and Clinical Presentation

Participants were recruited from a tertiary child and adolescent psychiatry outpatient clinic. Children were either referred by family physicians, primary care services, pediatric outpatient clinics, or other hospital departments, or were brought directly by their parents because of concerns about delayed speech and language development. The main reasons for presentation included delayed production of first words or sentences, limited expressive or receptive language abilities, reduced speech intelligibility, and parental or clinician concerns regarding co-occurring developmental or behavioral difficulties. Because the study was conducted in a tertiary referral setting, the study population may have included children with more complex developmental and medical histories than children with isolated speech and language difficulties managed in primary care or community settings.

Diagnostic Assessment

All children underwent a comprehensive clinical evaluation by a child and adolescent psychiatrist. Language disorder was diagnosed according to the criteria of the diagnostic and statistical manual of mental disorders, fifth edition.¹⁴ The assessment included a detailed developmental, medical, neurological, and family history obtained from the parent or legal guardian; direct observation of the child; and clinical evaluation of receptive and expressive language abilities, social communication, play skills, adaptive functioning, and overall developmental level.

Available speech-language assessment reports and developmental evaluations were reviewed. Children were referred to a speech and language therapist for standardized assessment when clinically indicated. Children were assessed by clinical examination for Autism Spectrum Disorder, Intellectual Disability, global developmental delay, and other psychiatric or neurodevelopmental conditions that could better explain the language difficulties; children with such conditions were excluded.

Hearing status was evaluated on the basis of parental history, medical records, previous hearing-screening or audiological-examination results, and clinical findings. Children with known hearing impairment or suspected hearing loss were referred for formal audiological assessment and excluded if the hearing impairment was considered sufficient to explain their speech and language difficulties. Developmental screening or standardized developmental assessment was performed when clinically indicated, particularly for children suspected of delays in developmental domains other than language.

Socio-Demographic Data Form

A brief data collection form was developed for this study. The form included items assessing the participants' ages, sex, number of siblings, family structure, and birth-related characteristics. It also included items assessing daily screen duration, age at first screen exposure, and related screen-use characteristics. No a priori power analysis was performed because the study was exploratory.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, version 27.0 (IBM Corp., Armonk, NY, USA). Given the study's descriptive and exploratory design, descriptive statistics were used as the primary analytical approach. Continuous variables were summarized as mean $\pm$ standard deviation, and categorical variables were presented as frequencies and percentages. For exploratory subgroup comparisons based on daily screen time (<1 hour/day vs. ≥ 1 hour/day) and age at first screen exposure (<12 months vs. ≥ 12 months), continuous variables were compared using the Independent Samples t-test or Mann-Whitney U test, as appropriate, whereas categorical variables were compared using the chi-square test or Fisher's exact test. Because of the small sample and subgroup sizes, these analyses were considered exploratory and hypothesis-generating, with a high-risk of type II error. A p-value of <0.05 was considered statistically significant.

RESULTS

A total of 43 children were included in the study, of whom 27 (62.8%) were male. The mean age was 47.18 ± 15.22 months, ranging from 24 to 72 months. The mean number of children per family was 2.04 ± 0.89 . Demographic, parental, family, and socio-economic characteristics are summarized in Table 1.

The mean gestational age was 36.67 ± 2.74 weeks, and the mean birth weight was 3068 ± 727 g. Cesarean delivery was reported in 31 children (72.1%); 14 (32.6%) required incubator care, and 10 (23.3%) had a history of postnatal hypoxia. Additional perinatal and postnatal characteristics are presented in Table 2.

Predominantly expressive impairment was the most frequent clinical profile, observed in 24 children (55.8%), while 14 children (32.6%) had mixed receptive-expressive impairment. Moderate impairment was the most common severity category, accounting for 21 children (48.8%). The mean duration of language-related concerns was 18.6 ± 9.4 months, and the mean age at first parental concern was 27.8 ± 8.7 months. Seventeen children (39.5%) had previously received speech and language therapy, while 23 children (53.5%) had not received any previous intervention. Further clinical characteristics are presented in Table 3.

With respect to screen exposure, 27 children (62.8%) had daily screen durations of at least 1 hour, including 12 (27.9%) who had 2 hours or more of exposure. Fourteen children (32.6%) were first exposed to screens before 12 months of age, while 32 children (74.4%) were first exposed before 24 months of age. Detailed screen exposure characteristics are presented in Table 4.

In exploratory subgroup analyses, no statistically significant differences were observed between children with daily screen exposure of <1 hour and those with exposure of ≥ 1 hour, or between children first exposed to screens before 12 months and those first exposed at 12 months or later (Table 5). Given the small subgroup sizes and limited statistical power,

Table 1. Demographic, family, and parental characteristics

Variable	n	%	Mean $\pm$ SD
Sex			
Male	27	62.8	-
Female	16	37.2	-
Mean age (months)			
Male	-	-	45.21 ± 15.93
Female	-	-	50.32 ± 13.81
Overall	-	-	47.18 ± 15.22
Mean number of children in the family			
Mean maternal age (years)	-	-	35.71 ± 5.42
Mean paternal age (years)	-	-	38.16 ± 4.49
Maternal education			
Primary school	11	25.6	-
Middle school	5	11.6	-
High school	9	20.9	-
University/college	18	41.9	-
Paternal education			
Primary school	3	7.0	-
Middle school	12	27.9	-
High school	13	30.2	-
University/college	15	34.9	-
Maternal employment status			
Employed	13	30.2	-
Unemployed	30	69.8	-
Paternal employment status			
Employed	36	83.7	-
Unemployed	7	16.3	-
Maternal medical illness			
Present	13	30.2	-
Absent	30	69.8	-
Paternal medical illness			
Present	10	23.3	-
Absent	33	76.7	-
Maternal psychiatric illness			
Present	1	2.3	-
Absent	42	97.7	-
Paternal psychiatric illness			
Present	2	4.7	-
Absent	41	95.3	-
Parental cohabitation status			
Living together	36	83.7	-
Separated	7	16.3	-
Income status			
Poor	11	25.6	-
Moderate	20	46.5	-
Good	12	27.9	-

SD: Standard deviation.

Table 2. Perinatal and postnatal characteristics

Variable	n	%	Mean $\pm$ SD
Gestational age (weeks)	-	-	36.67 $\pm$ 2.74
Birth weight (g)	-	-	3068 $\pm$ 727
Mode of delivery			
NSVD	12	27.9	-
Cesarean section	31	72.1	-
Incubator care			
Yes	14	32.6	-
No	29	67.4	-
Hypoxia			
Present	10	23.3	-
Absent	33	76.7	-
Prolonged jaundice			
Present	4	9.3	-
Absent	39	90.7	-

SD: Standard deviation, NSVD: Normal Spontaneous Vaginal Delivery.

Table 3. Clinical characteristics of speech and language disorder

Variable	n (%) or Mean $\pm$ SD
Language impairment profile	
Predominantly expressive	24 (55.8)
Predominantly receptive	5 (11.6)
Mixed receptive-expressive	14 (32.6)
Severity of language impairment	
Mild	12 (27.9)
Moderate	21 (48.8)
Severe	10 (23.3)
Duration of language concerns, months	18.6 $\pm$ 9.4
Age at first concern, months	27.8 $\pm$ 8.7
Previous interventions	
Speech-language therapy	17 (39.5)
Special education/developmental support	9 (20.9)
No previous intervention	23 (53.5)
Standardized language assessment	15 (34.9)
Formal audiological assessment	28 (65.1)

SD: Standard deviation.

these findings should not be interpreted as evidence of equivalence or as evidence that no differences existed between the groups.

DISCUSSION

Neurodevelopmental disorders arising in childhood are generally more common in males than in females. Several studies have also shown that delays in language development occur more frequently in boys.¹⁵⁻¹⁷ This has been attributed to the slower maturation of the central nervous system in males, as well as the proposed inhibitory effects of testosterone on the formation of appropriate neural connections.¹⁵⁻¹⁷ In the present sample, 62.8% of the participants were male. However, in the absence of a comparison group, this finding should not be interpreted as evidence of a sex-related difference in the occurrence of speech and language disorder.

Table 4. Screen exposure characteristics

Variable	n (%)
Daily screen time	
<1 hour	16 (37.2)
1-2 hours	15 (34.9)
2-4 hours	10 (23.3)
>4 hours	2 (4.7)
Age at first screen exposure	
0-6 months	5 (11.6)
6-12 months	9 (20.9)
12-18 months	12 (27.9)
18-24 months	6 (14.0)
>24 months	11 (25.6)

Table 5. Exploratory subgroup comparisons according to daily screen time and age at first screen exposure

Daily screen time: <1 hour/day vs. $\geq$1 hour/day			
Variable	<1 hour/day (n=16)	$\geq$ 1 hour/day (n=27)	p-value
Age, months, mean $\pm$ SD	43.8 $\pm$ 14.1	49.2 $\pm$ 15.6	0.284
Male sex, n (%)	8 (50.0)	19 (70.4)	0.182
Incubator care, n (%)	3 (18.8)	11 (40.7)	0.179
Hypoxia, n (%)	2 (12.5)	8 (29.6)	0.280
Age at first screen exposure: <12 months vs. $\geq$12 months			
Variable	<12 months (n=14)	$\geq$ 12 months (n=29)	p-value
Age, months, mean $\pm$ SD	46.2 $\pm$ 13.0	47.7 $\pm$ 16.4	0.756
Male sex, n (%)	11 (78.6)	16 (55.2)	0.129
Incubator care, n (%)	7 (50.0)	7 (24.1)	0.089
Hypoxia, n (%)	5 (35.7)	5 (17.2)	0.172

p-values were calculated using the Independent Samples t-test or Mann-Whitney U test for continuous variables, and the chi-square test or Fisher's exact test for categorical variables, as appropriate.

SD: Standard deviation.

Multiple factors associated with delayed language development have been identified, including hearing loss, recurrent middle ear infections, seizures, birth asphyxia, prematurity, low birth weight, and physical deformities affecting the head and neck region. Environmental and familial factors may also contribute to language difficulties. These include multilingualism in the family, a crowded family environment, parental history of speech and language difficulties, and low parental education.^{15,16,18}

In this cohort, nearly one-third of the children required incubator care after birth, and almost one-quarter had a history of postnatal hypoxia. These findings indicate the frequency of selected perinatal and postnatal medical characteristics in the study population. Because all participants had speech and language disorder, and no comparison group was included, the data cannot determine whether these characteristics were more frequent than in the general pediatric population or related to the development or severity of the disorder.

Among environmental factors, screen exposure has received considerable attention. Previous studies have discussed the possibility that screen use may reduce opportunities for reciprocal verbal interaction and other

social activities relevant to language acquisition. However, the present study did not evaluate these mechanisms and could not determine whether they applied to the children in this sample. In the present cohort, 74.4% of the children were exposed to screens before 2 years of age. This proportion reflects the reported screen-exposure history of the study population and should not be interpreted comparatively because no typically developing control group was included. Early screen exposure has also been reported in studies of children diagnosed with speech and language disorder.^{19,20} Nevertheless, the current findings do not demonstrate that early exposure was more frequent than it was in typically developing children, nor that it preceded or contributed to language difficulties.

Screen exposure should not be considered solely in terms of duration. Previous literature suggests that developmental outcomes reported in relation to screen use may vary according to the type of device, the educational or entertainment nature of the content, its developmental appropriateness, and the social context of use. For example, interactive and age-appropriate content accompanied by parental co-viewing and verbal engagement may provide a different developmental experience from passive, solitary viewing of rapidly changing entertainment content. Similarly, background television exposure may reduce the quantity and quality of parent-child interaction even when the child is not directly attending to the screen.¹⁰⁻¹² Therefore, children with similar daily screen durations may have substantially different patterns and contexts of exposure. Because the present study recorded only daily duration and age at first exposure, it provides a limited quantitative characterization and does not capture the qualitative or contextual dimensions of screen use.

Previous theoretical and empirical studies have suggested that young children may process rapidly changing audiovisual content differently from adults and may have difficulty integrating such stimuli with real-world experiences.^{21,22} However, the present study did not assess attention, cognitive processing, the ability to distinguish virtual from real content, or any proposed neurodevelopmental mechanism. No inferences regarding these mechanisms can, therefore, be drawn from the findings.

Chonchaiya and Pruksananonda²³ reported that children with delayed language development began watching television approximately 10 months before producing their first meaningful words. In that study, exposure to television for more than 2 hours per day combined with introduction of television before 12 months of age was associated with a higher reported likelihood of language delay.²³ These results are specific to the population and design of that study and should not be directly extrapolated to the current sample.

Perdana et al.²⁴ reported that both typically developing children and children with delayed language development were commonly exposed to television before 2 years of age. Children with language delay had longer reported viewing durations, although the between-group difference was not statistically significant. The authors also reported that children with language delay were more frequently left alone to watch television.²⁴ Although these findings provide relevant context, they were derived from a separate comparative study and therefore cannot be directly compared with the current results.

Previous studies have also examined characteristics such as prolonged television viewing, the presence of a television in the child's room,

viewing content in a language not used in the daily life, and the absence of parental accompaniment.²⁵ However, these characteristics were not assessed in the present study. Their frequency and potential relevance in this population therefore remain unknown.

In 2022, Zhang et al.²⁶ examined the relationship between screen time and cognitive developmental domains in preschool children. Regardless of device type, time spent in front of screens was not associated with expressive vocabulary, but was linked to working memory difficulties. The authors suggested that the absence of an association with vocabulary might reflect differences in the sensitivity of expressive and receptive language to television exposure during early childhood. They also noted that children's expressive vocabulary may be more limited than their receptive language capacity.²⁶ These findings further illustrate the heterogeneity of the existing literature and should be interpreted in relation to the specific outcomes and methods evaluated in each study.

Speech and language disorder is a heterogeneous clinical condition that may arise from a broad etiological spectrum, including neurodevelopmental disorders, environmental influences, and psychosocial factors. Comprehensive and systematic evaluation is therefore essential for accurate diagnosis and appropriate intervention.^{15,16} Early multidisciplinary assessment involving pediatrics, child and adolescent psychiatry, speech and language therapy, and, when indicated, neurology may help identify comorbid psychopathologies and developmental vulnerabilities. This approach not only facilitates accurate differential diagnosis but also enables the development of individualized and evidence-based treatment plans. Given the potential of early and targeted interventions to improve long-term developmental, communicative, and functional outcomes, timely referral and coordinated care are particularly important for children presenting with speech and language difficulties.⁴⁻⁷ These clinical recommendations are based on the broader literature and not on comparative or causal findings from the present study.

From a clinical perspective, the findings support the inclusion of a structured assessment of digital media habits as part of the comprehensive developmental evaluation of children presenting with speech and language difficulties. Such an assessment may include daily screen duration, age at first exposure, device type, content characteristics, parental co-viewing, and the social context of screen use. However, the evaluation of screen habits should not replace a broader assessment of hearing, developmental level, family characteristics, medical history, and neurodevelopmental comorbidities. This clinical recommendation reflects the importance of obtaining a comprehensive developmental history and should not be interpreted as evidence that screen exposure caused the language difficulties observed in the present sample.

Recent systematic reviews and meta-analyses indicate that associations reported between screen media exposure and early development are heterogeneous and depend substantially on the context of use. A 2023 meta-analysis examining screen media exposure and vocabulary development reported that the direction and magnitude of findings varied according to the type of media experience and learning context. More recent reviews have emphasized that age-appropriate educational content, active caregiver involvement, and interactive use may differ substantially from passive or solitary viewing in their reported relationships with language and cognitive outcomes. Conversely, background television exposure, age-inappropriate content, and reduced parent-child interaction during screen use have been

associated with less favorable developmental outcomes. These findings support the use of multidimensional screen-exposure assessments in future research, but do not permit conclusions about the developmental implications of screen use in the current population.¹⁰⁻¹³

The exploratory subgroup comparisons did not identify statistically significant differences. However, the small sample size and limited numbers within the comparison categories resulted in low statistical power and a high-risk of type II error. Therefore, the absence of statistically significant findings should not be interpreted as evidence of equivalence between the groups or as confirmation that no differences were present. These analyses were performed solely for hypothesis-generating purposes and cannot support definitive conclusions.

The contribution of this study lies in its structured characterization of screen-exposure patterns and selected socio-demographic, perinatal, and clinical features among children evaluated in a tertiary child psychiatry clinic. Because no comparison group was included, the observed frequencies cannot be interpreted as higher or lower than those observed in typically developing children. Likewise, the findings do not indicate that any of the recorded characteristics are associated with the presence, development, or severity of speech and language disorder. Nevertheless, they may help identify variables for investigation in future adequately powered case-control and longitudinal studies.

Study Limitations

This study has several limitations. First, the relatively small sample size and small screen-exposure subgroups resulted in low statistical power and a high-risk of type II error. Accordingly, the absence of statistically significant differences should not be interpreted as evidence of equivalence or as evidence that no relationship exists. The exploratory subgroup analyses were performed solely for hypothesis-generating purposes and cannot support definitive conclusions.

Second, all participants were recruited from a single tertiary child and adolescent psychiatry clinic. Children evaluated in such a setting may have more severe, complex, or multidimensional developmental and medical concerns than those with isolated speech and language difficulties managed in primary care or community settings. Referral and selection biases may therefore have influenced the observed frequencies of perinatal complications, clinical characteristics, and screen-exposure patterns. Consequently, the findings should not be interpreted as prevalence estimates and may not be generalizable to all children with speech and language disorders.

Third, the absence of a typically developing comparison group represents a major methodological limitation. The study cannot determine whether the observed screen-exposure patterns were more frequent, occurred earlier, or more prolonged than those in the general pediatric population. In addition, the descriptive cross-sectional design does not permit the assessment of temporal directionality or causal relationships. Thus, the study cannot determine whether screen exposure preceded, contributed to, followed, or was unrelated to the emergence of speech and language difficulties.

Fourth, screen-exposure data were obtained retrospectively from parental reports and may have been affected by recall bias, reporting inaccuracies, and social desirability bias. Parents may also have retrospectively attributed screen exposure to their child's language

difficulties, potentially influencing the reported exposure history. Furthermore, screen exposure was assessed only in terms of daily duration and age at first exposure. Important qualitative and contextual characteristics-including device type, educational versus entertainment content, developmental appropriateness, parental co-viewing, background television exposure, and interactive versus passive use-were not systematically evaluated. Because children with similar total screen durations may have substantially different media-use patterns and contexts, the measures used provide only a limited quantitative representation of screen exposure. The use of parental reports rather than a validated multidimensional media-use questionnaire or objective device-use data may also have introduced measurement bias.

Finally, standardized speech-language, developmental, and audiological assessments were not applied uniformly to all participants. Although children with autism spectrum disorder, intellectual disability, and hearing impairment sufficient to explain the language difficulties were excluded, variability in the diagnostic assessments may have introduced clinical heterogeneity. Other unmeasured developmental, familial, medical, and environmental factors may also have influenced the characteristics observed in the sample. These methodological limitations should be addressed in future research.

CONCLUSION

This descriptive study characterized screen exposure patterns and selected sociodemographic, perinatal, and clinical features in children aged 2-6 years diagnosed with speech and language disorder. Early screen exposure was frequently reported, with 74.4% of children first exposed before 24 months of age and 62.8% having at least 1 hour of daily screen exposure. These observations emphasize the relevance of documenting digital media habits as part of a comprehensive developmental assessment of children presenting with speech and language difficulties. However, the absence of a typically developing control group and the cross-sectional design preclude conclusions regarding associations, temporal relationships, or causality. Future prospective, controlled studies incorporating standardized language assessments and multidimensional measures of screen exposure are needed to clarify the potential relationship between digital media use and speech and language disorders.

MAIN POINTS

- This study aimed to characterize screen exposure patterns and selected sociodemographic, perinatal, and clinical features of children aged 2-6 years diagnosed with speech and language disorder.
- In this tertiary clinical sample, 74.4% of the children had first been exposed to screens before 2 years of age, and 62.8% had a reported daily screen duration of at least 1 hour. These findings are descriptive and should not be interpreted as indicating higher values than those observed in typically developing children.
- The findings highlight the importance of documenting screen exposure alongside clinical and developmental characteristics in children with speech and language disorders, while further controlled studies are needed to clarify potential associations.

ETHICS

Ethics Committee Approval: The study was initiated after obtaining approval from the Karabük University Non-Interventional Clinical Research Ethics Committee (decision no. 2025/2509, date: 30.10.2025).

Informed Consent: Written informed consent was obtained from the parents or legal guardians of all participants prior to inclusion in the study.

Footnotes

Authorship Contributions

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

DISCLOSURES

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

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

Declaration on the Use of Artificial Intelligence (AI): No artificial intelligence tools were used in the preparation of this manuscript.

REFERENCES

- Radesky JS, Christakis DA. Increased screen time: implications for early childhood development and behavior. *Pediatr Clin North Am.* 2016; 63(5): 827-39.
- Reid Chasiakos YL, Radesky J, Christakis D, Moreno MA, Cross C. Children and adolescents and digital media. *Pediatrics.* 2016; 138(5): e20162593.
- Kaur N, Gupta M, Malhi P, Grover S. Screen time in under-five children. *Indian Pediatr.* 2019; 56(9): 773-88.
- Matte-Landry A, Boivin M, Tanguay-Garneau L, Mimeau C, Brendgen M, Vitaro F, et al. Children with persistent versus transient early language delay: language, academic, and psychosocial outcomes in elementary school. *J Speech Lang Hear Res.* 2020; 63(11): 3760-74.
- Tinoco Mendoza G, Stack J, Abdel-Latif ME, Raman S, Garg P. Language outcomes at 4 years of linguistically diverse children born very preterm: an Australian retrospective single-centre study. *BMJ Paediatr Open.* 2023; 7(1): e001814.
- Ghosh E, Banerjee R. Early intervention strategies for language and literacy development in young dual language learners: a literature review. *Education Sciences.* 2025; 15(12): 1692.
- Dale PS, Price TS, Bishop DVM, Plomin R. Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *J Speech Lang Hear Res.* 2003; 46(3): 544-60.
- Carson L, Baker E, Munro N. A systematic review of interventions for late talkers: intervention approaches, elements, and vocabulary outcomes. *Am J Speech Lang Pathol.* 2022; 31(6): 2861-74.
- Pak NS, Chow JC, Dillehay KM, Kaiser AP. Long-term effects of early communication interventions: a systematic review and meta-analysis. *J Speech Lang Hear Res.* 2023; 66(8): 2884-99.
- Massaroni V, Delle Donne V, Marra C, Arcangeli V, Chieffo DPR. How screen time affects language development in early childhood: a systematic review. *Children(Basel).* 2023; 10(2): 282.
- Jing M, Ye T, Kirkorian HL, Mares ML. Screen media exposure and young children's vocabulary learning and development: a meta-analysis. *Child Dev.* 2023; 94(5): 1398-418.
- Mallawaarachchi S, Burley J, Mavilidi M, Howard SJ, Straker L, Kervin L, et al. Early childhood screen use contexts and cognitive and psychosocial outcomes: a systematic review and meta-analysis. *JAMA Pediatr.* 2024; 178(10): 1017-26.
- Bal M, Kara Aydemir AG, Tepetaş Cengiz GŞ, Altındağ A. Examining the relationship between language development, executive function, and screen time: a systematic review. *PLoS One.* 2024; 19(12): e0314540.
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Arlington, VA: American Psychiatric Association; 2013.
- Meschino WS. The child with developmental delay: an approach to etiology. *Paediatr Child Health.* 2003; 8(1): 16-9.
- Saeed HT, Abdulaziz B, Al-Daboon SJ. Prevalence and risk factors of primary speech and language delay in children less than seven years of age. *J Community Med Health Educ.* 2018; 8(2): 608.
- Chilosi AM, Brovedani P, Cipriani P, Casalini C. Sex differences in early language delay and in developmental language disorder. *J Neurosci Res.* 2023; 101(5): 654-67.
- Sansavini A, Favilla ME, Guasti MT, Marini A, Millepiedi S, Di Martino MV, et al. Developmental language disorder: early predictors, age for the diagnosis, and diagnostic tools. A scoping review. *Brain Sci.* 2021; 11(5): 654.
- Vandewater EA, Rideout VJ, Wartella EA, Huang X, Lee JH, Shim MS. Digital childhood: electronic media and technology use among infants, toddlers, and preschoolers. *Pediatrics.* 2007; 119(5): e1006-15.
- Collet M, Gagnière B, Rousseau C, Chapron A, Fiquet L, Certain C. Case-control study found that primary language disorders were associated with screen exposure. *Acta Paediatr.* 2019; 108(6): 1103-9.
- Samaniego CM, Pascual AC. The teaching and learning of values through television. *Int Rev Educ.* 2007; 53(1): 5-21.
- Büyükbaykal G. The effects of television on children. *Istanbul University Faculty of Communication Journal.* 2007; 28: 31-44.
- Chonchaiya W, Pruksananonda C. Television viewing associates with delayed language development. *Acta Paediatr.* 2008; 97(7): 977-82.
- Perdana SA, Medise BE, Purwaningsih EH. Duration of watching TV and child language development in young children. *Paediatr Indones.* 2017; 57(2): 99-103.
- Biedinger N. The influence of education and home environment on the cognitive outcomes of preschool children in Germany. *Child Dev Res.* 2011; (2).
- Zhang Z, Adamo KB, Ogden N, Goldfield GS, Okely AD, Kuzik N, et al. Associations between screen time and cognitive development in preschoolers. *Paediatr Child Health.* 2022; 27(2): 105-10.

The Epidemiological and Clinical Profile of the Patients Having Oocyte Donation Treatment in Northern Cyprus

✉ Burcu Özbakır

Department of Obstetrics and Gynecology, Near East University Faculty of Medicine; DESAM Research Institute, Near East University, Nicosia, Cyprus

Abstract

BACKGROUND/AIMS: Oocyte donation is a popular form of in vitro fertilisation (IVF) treatment for health tourists worldwide and in Northern Cyprus. Patients with various medical conditions requiring this treatment have diverse demographic profiles and clinical findings. Viewing patients holistically may provide new insights into treatment.

MATERIALS AND METHODS: This retrospective study included patients who underwent oocyte-donation IVF at the Near East University Hospital IVF Centre in Northern Cyprus between October 2021 and March 2024. The ages of female and male patients at the time of treatment, their smoking and alcohol habits, body mass index (BMI), endometrial parameters, transfer quality, sperm parameters, donor age and BMI, and oocyte and embryo evaluation parameters were assessed.

RESULTS: The mean ages of female and male patients were 41.91 ± 6.78 and 41.47 ± 9.21 , respectively, with most being overweight. Most patients preferred two-embryo transfers on day five of embryo development. While 18.1% of patients wished to add preimplantation genetic diagnosis to their treatment, 29.4% opted for sperm microchipping, and 11.3% chose sperm donation. The endometrial rim was thicker in younger female patients ($p < 0.001$) and in non-smoking female patients ($p = 0.021$). A trilaminar pattern of the endometrium was observed more frequently in younger ($p = 0.001$), non-smoking ($p = 0.031$), and non-alcohol-consuming ($p = 0.014$) patients.

CONCLUSION: Personalised treatments require evaluation of demographic and clinical variations among patients with diverse profiles and findings. This methodology may enable more effective modifications to clinical outcomes.

Keywords: Oocyte donation, ageing, smoking, endometrium, sperm count

INTRODUCTION

In vitro fertilisation (IVF) treatment with oocyte donation, first applied in 1984, has become a common practice in IVF clinics approved to perform oocyte donation.¹ The patient groups who require oocyte donation to achieve pregnancy are heterogeneous. The well-known aetiologies are women with poor ovarian reserve with premature ovarian insufficiency or age-related depletion of the ovarian reserve; advanced female age, women with genetically inherited disease or unidentified familial genetic condition; women with poor embryo or oocyte quantity and

quality; women with recurrent implantation failure and men with no partner or who have a trans female partner undergoing a gestational surrogacy program.² Given the range of needs for donor oocytes, it is no surprise that patient profiles can be quite diverse.

Oocyte donation programs exclude the ovarian factor. Thus, male and uterine parameters are essential. These programs provide favourable conditions for comparing the effects of demographic parameters with those of clinical parameters, such as sperm parameters and endometrial maturation. The standard semen evaluation in IVF procedures includes

To cite this article: Özbakır B. The epidemiological and clinical profile of the patients having oocyte donation treatment in Northern Cyprus. Cyprus J Med Sci. 2026;11(5):424-429

ORCID IDs of the authors: B.Ö. 0000-0002-0895-0487.



Corresponding author: Burcu Özbakır
E-mail: drburcuburcu@hotmail.com
ORCID ID: orcid.org/0000-0002-0895-0487

Received: 12.03.2026
Accepted: 02.10.2026
Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.
This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

semen volume, sperm concentration, motility, and morphology, or the total forward-motile sperm count.^{3,4} The standard endometrium evaluation in oocyte donation IVF procedures includes the measurement of the endometrial thickness.⁵ More recently, researchers have begun evaluating the trilinear pattern in endometrial assessment.⁶ We can also assess hormonal effects by examining cervical mucus.⁷

Objectives

Our study, which has two primary objectives, has significant implications for reproductive medicine. The first objective is to provide a comprehensive profile of patients undergoing oocyte-donation IVF treatment in Northern Cyprus, a health-tourism region with unique demographic and clinical characteristics. The second objective is to investigate the correlations between female factors [age, body mass index (BMI), lifestyle habits, and endometrial thickness] and male factors (age, BMI, lifestyle habits, and total forward-motile sperm count). These objectives are crucial for advancing our understanding of oocyte-donation IVF treatment.

MATERIALS AND METHODS

Study Design and Population

Our research was a retrospective cross-sectional study. We reviewed the medical records of women who participated in the oocyte-donation IVF treatment program at the Near East University Hospital IVF Centre between October 2021 and March 2024. We evaluated 320 embryo transfer cycles from 230 female patients who underwent the same endometrial preparation protocol.

Data Collection

We meticulously extracted the demographic data from the patient application forms, which were completed by the patient under the supervision of healthcare personnel, and the clinical and laboratory data from the medical files. We recorded add-on treatments, including preimplantation genetic diagnosis (PGD), sperm microchipping, and sperm donation, when present. We recorded the female patients' date of birth; month and year of embryo transfer; smoking status; alcohol consumption; weight and height at the time of embryo transfer; endometrial thickness and presence of a trilinear pattern at the time we started progesterone supplementation, or presence of cervical mucus at the time of embryo transfer, both of which indicate peak estrogen effect; transfer ease (easy, medium, hard); and presence of blood on the transfer catheter. We recorded the male patients' date of birth; month and year of embryo formation; smoking and alcohol consumption status; weight and height at the time of sperm sample deposition; and semen volume, sperm concentration, and sperm percentages for fast-forward and slow-forward motility. We also recorded the sperm donors' demographic and clinical information when donor sperm was used. We recorded the oocyte donors' dates of birth, weights, and heights at the time of donation. In the embryology laboratory, we recorded the number of metaphase-2 (M2) quality oocytes used; the number of 2 pronuclear (2 PN) formations at 16-18 hours after intracytoplasmic sperm injection (ICSI); the number of blastocyst formations on days 5-6 of embryo growth; the number of embryos transferred on days 5-6 of embryo growth; and whether the embryo transfer used fresh or frozen embryos. This meticulous data collection process ensures the reliability and accuracy of our study's findings.

We calculated the age of the female patient at the time of the embryo transfer, the age of the male patient or sperm donor at the time of the embryo formation, the age of the oocyte donor at the time of the donation, the BMI of the female and male patients as well as the oocyte donor, the total forward-motile sperm count of the semen sample used, the 2 PN formation percentage out of the M2 used, the blastocyst formation percentage out of the 2 PNs.

Ethics

Ethical approval was obtained from the Near East University Medical Faculty Ethical Committee (approval no: 2025/136, date: 25.09.2025).

Statistical Analysis

During the statistical analysis, we first examined the participants' demographic characteristics using frequency analysis. Before comparing any variables, we checked for normality by evaluating the "Skewness and Kurtosis values" ($-2 \leq \text{Skewness} \leq +2$; $-2 \leq \text{Kurtosis} \leq +2$). We used the Independent Samples t-test (t) to compare the variables when the data followed a normal distribution. For categorical variables, we used the Independent Samples chi-square test when expected cell counts exceeded 5. When the data were normally distributed, we evaluated the relationships between numerical variables using the "Pearson correlation test" (r = correlation coefficient). For descriptive statistics, we reported the number of observations (n), mean, standard deviation (SD), median, minimum, maximum, and significance value (p). Variables with a p-value < 0.25 in the bivariate analyses were considered candidates for the multiple linear regression model.⁸ Female age at embryo transfer and smoking status met this criterion and were entered simultaneously into the model, using the Enter method to evaluate their independent associations with endometrial thickness. We assessed the assumptions of multiple linear regression, including linearity; the normality and homoscedasticity of residuals; independence of errors; multicollinearity; and influential observations, before interpreting the model. We used a significance level of $p < 0.05$ for all calculations and interpretations. Statistical analyses were performed using the IBM SPSS, version 27.0 (IBM Inc., Armonk, NY, USA).

RESULTS

Sample Characteristics

Tables 1 and 2 present the characteristics of the research participants. This study, which included three hundred and twenty cycles, reported various participant attributes, such as smoking and alcohol consumption; embryo transfer ease; presence of blood on the embryo transfer catheter; number of transferred good-quality embryos; whether embryos were fresh or frozen; trilinear pattern/cervical mucus status; PGD; sperm microchipping; age and BMI of the female patient, the male patient, and the donors; total forward-motile sperm count; percentage of fast forward-motile sperm; number of M2 oocytes used; 2 PN number; 2 PN/M2 percentage; day 5-6 blastocyst number; and blastocyst/2 PN percentage.

We observed that male participants consumed almost twice as many cigarettes and four times as much alcohol as female participants. We observed that embryo transfer was easy for most participants, was moderate or difficult for 9.0% of participants, and that blood was present on the catheter in 15.6% of cases. We observed that frozen embryo transfers were slightly more frequent than fresh embryo transfers.

Table 1. Participants' demographic characteristics 1

Characteristic	Group	Number	%
Female-smoking existence	Smoking	78	24.4
	Not smoking	242	75.6
Female-alcohol consumption existence	Consuming	25	7.8
	Not consuming	295	92.2
Male-smoking existence	Smoking	136	42.5
	Not smoking	184	57.5
Male-alcohol consumption existence	Consuming	99	30.9
	Not consuming	221	69.1
Ease of the embryo transfer	Easy	288	90.0
	Medium	17	5.3
	Hard	15	4.7
Blood existence on the embryo transfer catheter	Not existing	270	84.4
	Existing	50	15.6
Transferred embryo number	1	100	31.3
	2	201	62.8
	3	19	5.9
Transferred embryo type	Fresh	131	40.9
	Frozen	189	59.1
Trilinear pattern of the endometrium/ cervical mucus	Not existing	119	37.2
	Existing	201	62.8
Preimplantation genetic diagnosis	Not added	262	81.9
	Added	58	18.1
Sperm microchipping	Not added	226	70.6
	Added	94	29.4
Sperm donation	Not added	284	88.7
	Added	36	11.3

Table 2. Participants' demographic characteristics 2

Characteristic	Mean	SD	Med	Min	Max
Female-age at ET	41.9	6.8	43	22.0	58.0
Female-BMI at ET	25.7	4.4	26	17.4	39.0
Male-age at embryo formation	41.5	9.2	42	19.0	72.0
Male-BMI at embryo formation	27.4	10.2	26	18.5	37.9
Male-total forward-motile sperm count	76.1	78.4	50	0.0	637.5
Male-a motile sperm percentage	6.3	7.4	5	0.0	40.0
Female-endometrial thickness at progesterone start date	9.0	1.6	9	5.8	16.0
Oocyte donor-age at donation	25.9	3.6	26	18.0	35.0
Oocyte donor-BMI at donation	21.8	3.7	21	16.7	33.8
M ² number reserved for the patient	15.0	6.0	15	1.0	38.0
2 PN number	11.0	4.3	11	1.0	27.0
2 PN/M ² percentage	76.4	16.8	78	12.5	100.0
Blastocyst number	6.5	3.2	7	1.0	18.0
Blastocyst/2 PN percentage	61.2	23.0	62	10.0	100.0

ET: Embryo transfer, SD: Standard deviation, Med: Median, Min: Minimum, Max: Maximum, BMI: Body mass index, M²: Metaphase-2 oocyte, 2 PN: 2 pronuclear formation.

Additionally, we reported that 31.3% of participants had a single embryo transferred, 62.8% had two, and 5.9% had three; decisions were based on patient preferences. With trilinear pattern and cervical mucus status reflecting estrogen levels in clinical practice, 62.81% exhibited a trilinear pattern or cervical mucus formation. When we examined the use of PGD, sperm microchipping, and sperm donation, the rates of participants benefiting from these additional treatments were 18.1%, 29.4%, and 11.3%, respectively.

The mean female age at the time of the embryo transfer was 41.9±6.8 (SD) years, and the mean BMI was 25.7±4.4 (SD) kg/m². The mean age of males at the time of embryo formation was similar to that of females. The mean for total forward-motile sperm count was 76.1±78.4 (SD) million/sample; for the fast-forward sperm percentage variable, the mean was 6.3±7.4 (SD) percent. The mean endometrial thickness among embryo transfer patients in the study was 9.0±1.6 mm, and the mean number of M2 eggs received by patients was 15.0±6.0. Considering the number of fertilised eggs (2 PN numbers) and the number of embryos surviving for five days, they were 11.0±4.3 and 6.5±3.2, respectively. The average donor age was 25.9±3.6 years, and the average donor BMI was within the normal range.

Demographic Factors and the Endometrium

The findings show that there is a weak negative correlation between endometrial thickness and women's age at embryo transfer, indicating that younger patients tend to have a thicker endometrium. On the other hand, are weak and non-significant relationships between endometrial thickness and BMI, and between BMI and women's age at embryo transfer (Table 3).

The mean endometrial thickness of non-smokers was 0.4 mm higher than that of smokers, a difference that was statistically significant. On the other hand, the mean endometrial thickness in women who did not consume alcohol was similar to that in women who consumed alcohol (Table 4).

There was no significant difference in women's BMI values based on the presence of a trilinear cervical mucus pattern. However, we found a substantial difference in the mean ages of women with and without a trilinear pattern of cervical mucus at the time of embryo transfer. The mean age of women with a trilinear pattern of cervical mucus was lower than that of women without this pattern (Table 5).

The study examined the association of smoking status and alcohol consumption status with the presence of a trilinear pattern in cervical mucus. The results indicated that smoking status was associated with the presence of a trilinear pattern of cervical mucus ($\chi^2=4.638$, $p=0.031$). Specifically, women who smoke (52.6%) are less likely to have a trilinear pattern of cervical mucus than those who do not smoke (66.1%). Additionally, the analysis revealed a statistically significant relationship between alcohol consumption status and the presence of a trilinear pattern of cervical mucus ($\chi^2=6.042$, $p=0.014$). Women who consumed alcohol (40.0%) were less likely to have a trilinear cervical mucus pattern than women who did not consume alcohol (64.7%).

As shown in Table 6, we used multiple linear regression to evaluate the independent associations of female age at embryo transfer and of smoking status with endometrial thickness. The overall model was statistically significant and explained 5.2% of the variance in endometrial thickness ($p<0.001$).

After adjustment for smoking status, the age of the female at embryo transfer remained independently and negatively associated with endometrial thickness. After adjustment for age, smoking status was not independently associated with endometrial thickness.

Demographic Factors and the Sperm

Correlation analysis indicated that the relationship between total forward-motile sperm count and BMI was weak and not statistically significant ($r=-0.032$, $p=0.568$). There was also a weak and statistically non-significant relationship between the total forward-motile sperm count and the male age ($r=-0.071$, $p=0.208$). The relationship between male age and BMI during embryo formation was also weak and not statistically significant ($r=-0.023$, $p=0.687$).

Regarding male smoking, we calculated the mean total forward-motile sperm count for non-smokers as $73.9 \pm 71.71.1$ (SD) million/sample, and for smokers as 79.1 ± 87.5 (SD) million/sample. We found no significant difference in total forward-motile sperm counts between groups ($t=-0.594$, $p=0.556$). In terms of alcohol use in male patients, the mean total forward-motile sperm count of those who did not consume alcohol was calculated as 73.4 ± 74.6 (SD) million/sample, and the

mean total forward-motile sperm count of those who drank alcohol was 82.2 ± 86.3 (SD) million/sample. The results showed no significant difference in total forward-motile sperm count between the groups ($t=-0.928$, $p=0.381$). These findings indicate that male patients' smoking or alcohol-use status did not have a statistically significant effect on the total forward-motile sperm counts.

DISCUSSION

Main Findings

Demographic factors

Some studies examined only specific demographic characteristics of oocyte donation patients. The first study grouped the women into age categories: <40 or ≥ 40 .⁹ The study was interested only in information about the patients' previous pregnancies and IVF. It provided no information about the clinical aspects or the male patient. The second study also grouped the female patients into four age groups. It also included previous antenatal history, maternal education, and marital history. This study also included semen abnormalities.¹⁰ Neither study reported personal clinical parameters such as BMI, endometrial status, or laboratory follow-up. We added these parameters and did not group patients by age, instead treating age as a continuous variable to assess the potential effects of other parameters. Another study examined the relationship between BMI and pregnancy outcomes and found no correlation.¹¹ Our research found no correlation between the female patients' BMI and endometrial maturity.

Endometrium

A study evaluated the impact of age on the endometrium in patients undergoing oocyte donation.¹² They grouped the patients by age and analysed clinical and pregnancy outcomes. The research concluded that age did not affect endometrial thickness but reduced the chances of pregnancy. Another more recent review that sought similar information suggested a slight reduction in pregnancy rates among advanced-age patients, even with oocyte donation, but did not assess endometrial thickness.¹³ A recent study of the ultrasound appearance of the endometrium evaluated endometrial thickness and the presence of the trilinear pattern and found no correlation between them. It also did not compare age with endometrium; instead, it focused only on pregnancy rate, which was insignificant.⁶ A more recent review of the biological basis of endometrial receptivity revealed an age-related decline. Still, they did not search for this information among oocyte-donation patients receiving hormonal supplementation.¹⁴ A review analysing the outcomes of oocyte-donor patients revealed a decline in pregnancy success rates with increasing age. The study attempted to elucidate the potential causes of endometrial ageing.¹⁵ Our study used multiple linear regression analysis and found that age significantly affected endometrial thickness after adjustment for smoking. A meta-analysis evaluating the cervical mucus in fertile women also revealed

Table 3. The relationship between the endometrial thickness and the characteristics of the female patients

Variable	Value	Endometrial thickness	Female age at the time of ET	Female patient's BMI
Endometrial thickness	r p	1		
Female age at the time of ET	r p	-0.204 <0.001	1	
Female patient's BMI	r p	0.053 0.343	0.096 0.087	1

It is the p-value of the change in r value of endometrial thickness with increasing female age at the time of embryo transfer.
 r : Correlation coefficient, p : Value of significance, ET: Embryo transfer, BMI: Body mass index.

Table 4. The relationship between smoking or alcohol consumption parameters and the endometrial thickness of female patients

Variable	Group	Mean	SD	Med	t	p
Female-smoking	Not smoking	9.1	1.6	9	2.31	0.021
	Smoking	8.7	1.3	8		
Female-alcohol consumption	Not consuming	9.0	1.5	9	0.04	0.98
	Consuming	9.0	2.1	8		

It is the p-value of the difference between smoking and non-smoking female participants in endometrial thickness.
SD: Standard deviation, Med: Median, t: Independent Samples t-test, p: Value of significance.

Table 5. The relationship between the female age at the time of embryo transfer and BMI, and the existence of the trilinear pattern or cervical mucus

Trilinear pattern/cervical mucus	Not existing			Existing			t	p
	Mean	SD	Med	Mean	SD	Med		
Female-BMI at ET	25.7	4.1	25.7	25.8	4.6	25.7	-0.15	0.89
Female-age at ET	43.6	5.1	44.0	40.9	7.4	43.0	3.50	0.001

It is the p-value of the difference between trilinear pattern existence and non-existence with increasing female age at the time of embryo transfer.
BMI: Body mass index, ET: Embryo transfer, SD: Standard deviation, Med: Median, t: Independent Samples t-test, p: Value of significance.

Table 6. Multiple linear regression analysis of factors associated with endometrial thickness

Variable	B	SE	β	t	p	95% CI
Constant	11.342	0.572	-	19.839	<0.001	10.217, 12.467
Female age at embryo transfer, years	-0.044	0.013	-0.19	-3.452	0.001	-0.070, -0.019
Smoking status						
Non-smoker (ref)	-	-	-	-	-	-
Smoker	-0.381	0.203	-0.104	-1.879	0.061	-0.781, 0.018
Dependent variable: endometrial thickness. The non-smoker group was used as the reference category for smoking status. Model: $R=0.228$, $R^2=0.052$, adjusted $R^2=0.046$, Durbin-Watson =1.974. B: Unstandardized regression coefficient, SE: Standard error, β : Standardised regression coefficient, CI: Confidence interval.						

a decrease in cervical mucus maturity with advanced age.⁷ Our study also observed that advanced age had a mild negative impact on the endometrium and cervical mucus.

Sperm

There was research evaluating the relationship between semen analysis and paternal age over time, but they did not find any significant correlation.^{16,17} The relationship between BMI and total motile sperm count showed conflicting results across studies. Some found a significant negative correlation between advanced paternal age and the motile sperm count, whereas others found no correlation.¹⁸⁻²¹ The smoking and semen parameters, especially sperm motility, also yielded conflicting results, with positive, negative or no impact.²²⁻²⁵ The relationship between alcohol consumption and total motile sperm count showed conflicting results as well.²⁶⁻²⁸ In our research, the semen analysis, in particular total motile sperm count, was found not to be affected by paternal age, BMI, smoking, or alcohol.

Sturdy Limitations

One of the strengths of the research was the inclusion of demographic factors, such as biological age, BMI, smoking, and alcohol consumption habits, in the evaluation of endometrial maturation in female patients. However, a limitation was that we did not evaluate these parameters with respect to pregnancy achievement; rather, we only evaluated endometrial thickness and maturation. A similar pattern was observed among male patients. We included the effects of biological age, BMI, smoking, and alcohol consumption on the total forward-motile sperm count, an approach that was relatively novel in the field. We did not evaluate these parameters in relation to pregnancy outcomes. Another limitation was that we did not group the patients by smoking status and alcohol consumption levels when assessing the total forward motile sperm count or the endometrial maturation. If we had done so and had the dose-response data, we could have shown the effect of smoking or alcohol consumption more precisely.

Another strength was the inclusion of the total forward motile sperm count as an outcome, because sperm used for ICSI are preferably forward motile and the embryologist's ability to select sperm that meet the chosen criteria depends on sperm count and concentration, which may affect the outcome. To date, the closest parameter evaluated was the total motile sperm count, which includes non-progressive sperm. This research is unique in this respect.

Another strength of the research was that it included more than 300 evaluation cycles, which gave us a more precise assessment. However,

it did not group patients by the aetiology underlying the need for oocyte donation, which could have improved our understanding of the demographic findings.

CONCLUSION

When we described the demographic features of the population undergoing oocyte donation IVF treatment, we found that most patients preferred transferring two embryos rather than one and did not prefer PGD. The embryo growth rates were consistent with the previous data.

Almost two-thirds of the female patients had signs of adequate endometrial preparation, with endometrial thickness >7.4 mm; notably, patients' age affected endometrial thickness and maturation. Almost one-quarter of the female patients were smokers, and smoking appeared to affect the endometrial thickness and maturation. Alcohol did not significantly affect the endometrial thickness but did appear to affect the endometrial maturation.

More than forty per cent of the male patients were smokers, and one-third consumed alcohol; however, we observed that both parameters had no significant effect on the total forward-motile sperm count. BMI did not affect endometrial thickness, endometrial maturation, or total forward-motile sperm count.

When evaluating success, we focus solely on the pregnancy rates, but sometimes overlook the preparatory steps for embryo transfer. These steps may shed light on a higher pregnancy rate and alter the results.

MAIN POINTS

- Patients undergoing oocyte donation treatment had their profiles evaluated comprehensively in a medical research study, and the results were noteworthy.
- The endometrial thickness appeared greater in younger, nonsmoking patients than in older patients.
- The endometrial trilinear pattern was seen more often among younger, non-smoking, and non-alcohol-consuming groups compared with the older, smoking, and alcohol-consuming groups.
- Neither male age nor body mass index seemed to affect the mean total forward-motile sperm count.
- The male patients' history of smoking or alcohol consumption did not seem to affect the mean total forward-motile sperm count.

ETHICS

Ethics Committee Approval: Ethical approval was obtained from the Near East University Medical Faculty Ethical Committee (approval no: 2025/136, date: 25.09.2025).

Informed Consent: The informed consents of the all participants/patients were obtained at the time of the treatment protocol.

DISCLOSURES

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

Declaration on the Use of Artificial Intelligence (AI): No artificial intelligence tools were used in the preparation of this manuscript.

REFERENCES

1. Lutjen P, Trounson A, Leeton J, Findlay J, Wood C, Renou P. The establishment and maintenance of pregnancy using in vitro fertilization and embryo donation in a patient with primary ovarian failure. *Nature*. 1984; 307(5947): 174-5.
2. Practice Committee of the American Society for Reproductive Medicine and the Practice Committee for the Society for Assisted Reproductive Technology. Electronic address: ASRM@asrm.org. Guidance regarding gamete and embryo donation. *Fertil Steril*. 2021; 115(6): 1395-410.
3. Capelouto SM, Nagy ZP, Shapiro DB, Archer SR, Ellis DP, Smith AK, et al. Impact of male partner characteristics and semen parameters on in vitro fertilization and obstetric outcomes in a frozen oocyte donor model. *Fertil Steril*. 2018; 110(5): 859-69.
4. Setti AS, Braga DPAF, Iaconelli Junior A, Borges Junior E. Increasing paternal age and ejaculatory abstinence length negatively influence the intracytoplasmic sperm injection outcomes from egg-sharing donation cycles. *Andrology*. 2020; 8(3): 594-601.
5. Glujovsky D, Pesce R, Sueldo C, Quinteiro Retamar AM, Hart RJ, Ciapponi A. Endometrial preparation for women undergoing embryo transfer with frozen embryos or embryos derived from donor oocytes. *Cochrane Database Syst Rev*. 2020; 10(10): CD006359.
6. Oluborode B, Burks H, Craig LB, Peck JD. Does the ultrasound appearance of the endometrium during treatment with assisted reproductive technologies influence pregnancy outcomes? *Hum Fertil (Camb)*. 2022; 25(1): 166-75.
7. Najmabadi S, Schliep KC, Simonsen SE, Porucznik CA, Egger MJ, Stanford JB. Cervical mucus patterns and the fertile window in women without known subfertility: a pooled analysis of three cohorts. *Hum Reprod*. 2021; 36(7): 1784-95.
8. Md Suhaimi TM, Ismail A, Ismail R, Rasudin NS, Mohd Noor N, Jayapalan A, et al. Influence of maternal risk perception and vaccination knowledge on childhood vaccination intentions. *BMC Public Health*. 2025; 25(1): 671.
9. Sauer MV, Paulson RJ. Demographic differences between younger and older recipients seeking oocyte donation. *J Assist Reprod Genet*. 1992; 9(4): 400-4.
10. Sauer MV, Ary BR, Paulson RJ. The demographic characterization of women participating in oocyte donation: a review of 300 consecutively performed cycles. *Int J Gynaecol Obstet*. 1994; 45(2): 147-51.
11. Jungheim ES, Schon SB, Schulte MB, DeUgarte DA, Fowler SA, Tuuli MG. IVF outcomes in obese donor oocyte recipients: a systematic review and meta-analysis. *Hum Reprod*. 2013; 28(10): 2720-7.
12. Flamigni C, Borini A, Violini F, Bianchi L, Serrao L. Oocyte donation: comparison between recipients from different age groups. *Hum Reprod*. 1993; 8(12): 2088-92.
13. Paulson RJ. Does the age of the recipient influence the probability of pregnancy among recipients of oocyte donation? *Fertil Steril*. 2014; 101(5): 1248-9.
14. Pathare ADS, Loid M, Saare M, Gidlöf SB, Zamani Esteki M, Acharya G, et al. Endometrial receptivity in women of advanced age: an underrated factor in infertility. *Hum Reprod Update*. 2023; 29(6): 773-93.
15. Melnick AP, Rosenwaks Z. Oocyte donation: insights gleaned and future challenges. *Fertil Steril*. 2018; 110(6): 988-93.
16. Legro RS, Wong IL, Paulson RJ, Lobo RA, Sauer MV. Recipient's age does not adversely affect pregnancy outcome after oocyte donation. *Am J Obstet Gynecol*. 1995; 172(1 Pt 1): 96-100.
17. Santiago J, Silva JV, Santos MAS, Fardilha M. Age-dependent alterations in semen parameters and human sperm MicroRNA profile. *Biomedicines*. 2023; 11(11): 2923.
18. Ma J, Wu L, Zhou Y, Zhang H, Xiong C, Peng Z, et al. Association between BMI and semen quality: an observational study of 3966 sperm donors. *Hum Reprod*. 2019; 34(1): 155-62.
19. Macdonald AA, Stewart AW, Farquhar CM. Body mass index in relation to semen quality and reproductive hormones in New Zealand men: a cross-sectional study in fertility clinics. *Hum Reprod*. 2013; 28(12): 3178-87.
20. Eisenberg ML, Kim S, Chen Z, Sundaram R, Schisterman EF, Buck Louis GM. The relationship between male BMI and waist circumference on semen quality: data from the LIFE study. *Hum Reprod*. 2014; 29(2): 193-200.
21. Service CA, Puri D, Al Azzawi S, Hsieh TC, Patel DP. The impact of obesity and metabolic health on male fertility: a systematic review. *Fertil Steril*. 2023; 120(6): 1098-111.
22. Tang Q, Pan F, Wu X, Nichols CE, Wang X, Xia Y, et al. Semen quality and cigarette smoking in a cohort of healthy fertile men. *Environ Epidemiol*. 2019; 3(4): e055.
23. Henriques MC, Santiago J, Patrício A, Herdeiro MT, Loureiro S, Fardilha M. Smoking induces a decline in semen quality and the activation of stress response pathways in sperm. *Antioxidants (Basel)*. 2023; 12(10): 1828.
24. Bundhun PK, Janoo G, Bhurtu A, Teeluck AR, Soogund MZS, Pursun M, et al. Tobacco smoking and semen quality in infertile males: a systematic review and meta-analysis. *BMC Public Health*. 2019; 19(1): 36.
25. Rehman R, Zahid N, Amjad S, Baig M, Gazzaz ZJ. Relationship between smoking habit and sperm parameters among patients attending an infertility clinic. *Front Physiol*. 2019; 10: 1356.
26. Ricci E, Al Beitawi S, Cipriani S, Candiani M, Chiaffarino F, Viganò P, et al. Semen quality and alcohol intake: a systematic review and meta-analysis. *Reprod Biomed Online*. 2017; 34(1): 38-47.
27. Bai S, Wan Y, Zong L, Li W, Xu X, Zhao Y, et al. Association of alcohol intake and semen parameters in men with primary and secondary infertility: a cross-sectional study. *Front Physiol*. 2020; 11: 566625.
28. Ricci E, Noli S, Ferreri S, La Vecchia I, Cipriani S, De Cosmi V, et al. Alcohol intake and semen variables: cross-sectional analysis of a prospective cohort study of men referring to an Italian Fertility Clinic. *Andrology*. 2018; 6(5): 690-6.

Global Trends and Emerging Research Themes in Patient-Reported and Patient-Centered Outcomes after Cataract Surgery: A Bibliometric and Science Mapping Analysis

✉ Dilber Keskinel¹, ✉ Gülcan Dürüst Sakallı²

¹Clinic of Ophthalmology, Dr. Burhan Nalbantoğlu State Hospital, Nicosia, Cyprus

²Department of Nursing, Eastern Mediterranean University Faculty of Health Sciences, Famagusta, Cyprus

Abstract

BACKGROUND/AIMS: Research on cataract surgery outcomes has increasingly expanded beyond visual acuity to include visual function, quality of life, patient satisfaction, and other patient-reported and patient-centered outcomes. This study aimed to examine global research trends and emerging themes in patient-reported and patient-centered outcomes following cataract surgery using bibliometric and science mapping methods.

MATERIALS AND METHODS: A literature search was conducted in the Web of Science Core Collection on 11 September 2026 using the topic field. Following document-type screening, 2,704 article and review article publications were analyzed using the bibliometrix package in R. Scientific production, citation characteristics, sources, authors, countries, institutions, collaboration patterns, keywords, and trend topics were evaluated.

RESULTS: The dataset comprised 2,704 publications from 472 sources published between 1980 and 2026. The mean number of citations per document was 26.42, the annual scientific production growth rate was 11.57%, and the international co-authorship rate was 22.15%. The United States was the most productive country (n=477); the University of London was the most productive institution (n=271); and Pesudovs K was the most productive author (n=53). Research themes progressed from visual outcomes and quality of life to visual function, patient satisfaction, spectacle independence, newer intraocular lens technologies, and patient-reported outcome measures. Psychometric properties and Rasch analysis were also prominent methodological themes.

CONCLUSION: Research on cataract surgery outcomes has demonstrated substantial growth and an expanding focus on patient-reported and patient-centered outcomes. Standardized patient-reported outcome measures and greater international, multicenter collaboration may strengthen future research in this field.

Keywords: Cataract extraction, quality of life, patient satisfaction, patient-reported outcome measures, bibliometric analysis

INTRODUCTION

Cataract remains one of the major causes of the global burden of vision loss, particularly in the context of population aging. According to global estimates for 2020, cataract was associated with approximately 17 million cases of blindness and 83.5 million cases of moderate-to-

severe vision impairment. Population growth and aging have also been identified as important contributors to the increasing global burden of cataract-related vision loss.¹ Cataract surgery is a primary treatment for cataract-related vision loss; however, postoperative visual outcomes may vary substantially across countries and regions.² Importantly, evaluating

To cite this article: Keskinel D, Dürüst Sakallı G. Global trends and emerging research themes in patient-reported and patient-centered outcomes after cataract surgery: a bibliometric and science mapping analysis. Cyprus J Med Sci. 2026;11(5):430-437

ORCID IDs of the authors: D.K. 0000-0002-3454-9902; G.D.S. 0000-0002-6361-6924.



Corresponding author: Dilber Keskinel
E-mail: dilber_keskinel@yahoo.com
ORCID ID: orcid.org/0000-0002-3454-9902

Received: 18.09.2026
Accepted: 12.10.2026
Publication Date: 16.10.2026



Copyright© 2026 The Author(s). Published by Galenos Publishing House on behalf of Cyprus Turkish Medical Association.
This is an open access article under the Creative Commons AttributionNonCommercial 4.0 International (CC BY-NC 4.0) License.

the success of cataract surgery solely through clinical indicators such as visual acuity may not fully capture patients' postoperative visual functioning, difficulties in daily activities, or changes in quality of life. Patient-reported assessments can provide complementary information to clinical measures by capturing patients' perceptions of visual function and the effects of surgery on their daily lives.^{3,4}

Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs) are increasingly used to evaluate the benefits of cataract surgery from the patient's perspective. These measures capture patients' own assessments of visual function, vision-related difficulties, symptoms, and the impact of treatment on their daily lives. In particular, instruments such as the Catquest-9SF have demonstrated good validity, reliability, and sensitivity for assessing changes in visual function following cataract surgery.³ The development of additional standardized PROMs for assessing visual symptoms and patients' visual experiences further indicates the expanding scope of patient-reported assessment in cataract research.⁴ The Cat-PROM5 and Catquest-9SF have also demonstrated strong psychometric performance in the assessment of cataract surgery outcomes.⁵ More recently, Catquest-9SF has been used to assess patient-reported visual function following cataract surgery in the United States Veterans Health Administration, supporting its continued applicability across diverse patient populations.⁶

The scope of PROs in cataract research extends beyond changes in visual function to include visual quality, symptoms, daily functioning, and satisfaction with treatment outcomes. A multicenter study found that patient-reported satisfaction following cataract surgery was associated with visual quality and aspects of the care process, while dysphotopsia was strongly associated with dissatisfaction.⁷ Similarly, standardized assessment of visual symptoms has highlighted patient-reported experiences that may not be fully explained by conventional clinical measures.⁴ Collectively, these findings support the importance of incorporating patients' experiences, functional outcomes, symptoms, and satisfaction alongside conventional clinical indicators when evaluating cataract surgery outcomes.

The increasing diversity of PROs, patient populations, and measurement approaches suggests that research on cataract surgery outcomes has developed into a multidimensional field. Beyond visual function and vision-related quality of life, increasing attention has been given to patient satisfaction, visual symptoms, daily functioning, and other outcomes reflecting patients' experiences. This broader perspective indicates a shift toward more comprehensive patient-centered outcome assessment.⁵⁻⁷ However, the researchers, institutions, countries, journals, and research themes shaping this expanding field, as well as their development over time, cannot be comprehensively characterized through individual clinical studies alone.

Bibliometric analysis and science mapping provide useful approaches for examining the structure and development of expanding research fields. Bibliometric methods enable evaluation of publication and citation patterns and identification of productive authors, institutions, countries, and journals, whereas science mapping examines relationships among research actors and concepts and identifies thematic developments within a field.^{8,9} These approaches can therefore provide an overview not only of the volume and distribution of scientific production but also of the themes that have shaped a research field and those that are emerging over time.

Despite the growing diversity of research on patient-reported and patient-centered outcomes after cataract surgery, the global structure and temporal development of this research field have not been comprehensively characterized. A comprehensive assessment of the researchers, institutions, countries, journals, collaboration patterns, and evolving research themes associated with different patient-reported and patient-centered outcome domains is needed. Such an assessment may help clarify the structure of the existing evidence base and identify established and emerging directions for future research.

Therefore, this study aimed to examine the global structure of the scientific literature on patient-reported and patient-centered outcomes after cataract surgery within the Web of Science (WoS) Core Collection, using bibliometric analysis and science-mapping methods. Specifically, we evaluated the temporal development of scientific production; identified the most productive authors, institutions, countries, and journals; examined publication and citation patterns; assessed scientific collaboration patterns; and identified established and emerging research themes in the field.

MATERIALS AND METHODS

Research Design

This study was designed as a bibliometric and science-mapping analysis to examine the global developments, research trends, and emerging themes in the literature on patient-reported and patient-centered outcomes after cataract surgery. Publication and citation characteristics, research productivity, collaboration patterns, and thematic developments were evaluated.

Data Source and Search Strategy

Bibliographic data were obtained from the WoS Core Collection. The literature search was conducted on 11 September 2026 using the topic (TS) field, which searches the title, abstract, author keywords, and keywords plus fields. The search strategy included terms related to cataract surgery, quality of life, visual function, patient satisfaction, patient experience, PROs, postoperative recovery, and daily functioning.

TS = ("cataract surgery") and TS = ("quality of life" or "health-related quality of life" or "vision-related quality of life" or "visual function" or "visual functioning" or "patient satisfaction" or "patient experience" or "patient-reported outcome" or "patient-reported outcomes" or "quality of recovery" or "postoperative recovery" or "daily functioning" or "activities of daily living").

Only records classified as article or review article were retained for analysis, resulting in a final dataset of 2,704 publications. Because WoS document-type classifications may overlap, eligibility was based on records classified as either article or review article.

To evaluate the precision of the search strategy and the potential inclusion of false-positive records, a random sample of 100 records from the final dataset was independently assessed by two reviewers based on their titles and abstracts. Records were evaluated for their relevance to cataract surgery and patient-reported or patient-centered outcomes. Initial agreement between the two reviewers was 98% (98/100 records). Disagreements concerning two records were resolved through discussion and consensus.

Inclusion and Exclusion Criteria

Records were included if they were classified as “article” or “review article” and if they addressed outcomes related to cataract surgery within the scope of the search strategy. Records not classified as article or review article were excluded. The final dataset comprised 2,704 publications.

Ethics

This study was a bibliometric analysis of bibliographic and citation data from the published scientific literature. No data were collected from human participants, and no personal data or individual patient records were collected. No intervention was performed. Therefore, the study did not involve human subjects research requiring ethics committee approval. The analysis was based exclusively on bibliographic records retrieved from the WoSCore Collection.

Statistical Analysis (Bibliometric Analysis and Science Mapping)

Bibliometric analyses were performed in R using the bibliometrix package. The analysis included publication and citation characteristics, annual scientific production, source distribution and Bradford’s law, author productivity and local citations, country-level production and international collaboration, institutional productivity, and keyword and trend-topic analyses. Country-level production was assessed using total publications, single-country publications (SCP), multiple-country publications (MCP), and international co-authorship rates. Institutional productivity was determined from institutional affiliations reported in WoS records. Author keywords and keywords plus were analyzed to determine keyword frequencies and their temporal distribution. Trend-topic analysis was performed using the Bibliometrix package to examine the temporal distribution of frequently occurring terms across publication years. The analysis was descriptive, based on term frequency and the temporal distribution of keywords within the dataset. In this study, the term “emerging” was used descriptively to refer to topics whose occurrence was concentrated in the most recent publication years, rather than to denote a statistically defined or inferentially tested category. No inferential threshold was applied to classify a topic as emerging. The annual scientific production growth rate was calculated by the bibliometrix package as the compound annual growth rate (CAGR) over the 1980-2026 study period, based on the number of publications in the first and last years of the dataset. CAGR was calculated as $[(N_{\text{final}}/N_{\text{initial}})^{(1/n)} - 1] \times 100$, where N_{initial} and N_{final} represent the numbers of publications in the first and last years, respectively, and n represents the number of years between these two time points. To account for differences in citation windows across publication years, a year-normalized citation score (Normalized TC) was also considered. In bibliometrix, documents are grouped according to publication year, and the total citation count of each document is normalized relative to the citation performance of documents published in the same year. Therefore, a Normalized TC value greater than 1 indicates citation performance above the average for the corresponding publication-year cohort. This approach was used to reduce the influence of unequal citation exposure times when comparing documents published in different years.

All bibliometric findings were analyzed and interpreted descriptively. No inferential statistical tests were performed; therefore, the observed patterns, trends, and associations should not be interpreted as statistically tested differences or relationships.

RESULTS

General Bibliometric Characteristics of the Dataset

A total of 2,704 documents published between 1980 and 2026 were analyzed. These documents were published in 472 sources and involved 9,787 authors. The average number of co-authors per document was 5.37, with a median of 5.00 [interquartile range (IQR): 3.00-7.00] and a range of 1-248. The international co-authorship rate was 22.15%, and the average number of citations per document was 26.42. Given the positively skewed citation distribution, the median citation count was 12.00 (IQR: 3.00-30.25), with a range of 0-1,254. The dataset contained 50,851 references, with an annual scientific production growth rate of 11.57%. The average document age was 10.5 years, with a median of 9.00 years (IQR: 4.00-16.00) and a range of 0-46 years. Detailed characteristics are presented in Table 1.

Annual Scientific Production

Scientific production demonstrated a clear long-term increasing trend between 1980 and 2026 (Figure 1). The first publication was recorded in 1980. Publication output increased from 33 documents in 2000 to 66 documents in 2010 and 137 documents in 2020. This increase continued in subsequent years, reaching 137 publications in 2021, 148 in 2022, 157 in 2023, 162 in 2024, and 194 in 2025, the highest annual output in the dataset. A total of 154 publications were recorded in 2026. Because the search was conducted on 11 September 2026, this value does not represent the complete calendar year.

Citation Performance by Year

Average citation performance varied substantially across publication years, with higher values among older publications and lower values among recent publications (Figure 2). The average total citations per document was 109.00 in 1993 and 91.15 in 1998, compared with 12.34 in 2022, 12.62 in 2023, 5.59 in 2024, 2.53 in 2025, and 0.51 in 2026. The lower citation counts among recent publications are likely related, at

Table 1. Basic bibliometric characteristics of the research dataset	
Bibliometric indicator	Result
Research period	1980-2026
Sources (journals, books, etc.)	472
Total number of documents	2.704
Annual scientific production growth rate	%11.57
Average document age	10.5 years
Median document age (IQR)	9.00 years (4.00-16.00)
Document age range	0-46 years
Average citations per document	26.42
Median citations per document (IQR)	12.00 (3.00-30.25)
Citation range	0-1.254
Total number of references	50.851
Number of authors	9.787
Number of single-authored documents	113
Average number of co-authors per document	5.37
Median number of co-authors per document (IQR)	5.00 (3.00-7.00)
Co-author range	1-248
International co-authorship rate	%22.15
IQR: Interquartile range.	

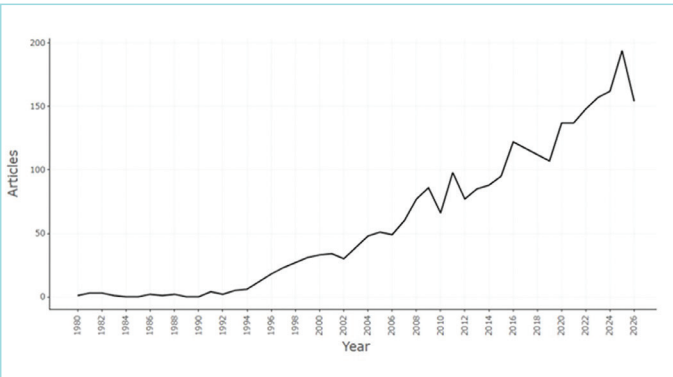


Figure 1. Annual scientific production from 1980 to 2026. Data for 2026 are partial and include publications indexed through 11 September 2026. The dashed segment indicates the incomplete observation period and should not be interpreted as a decline in annual scientific production.

least in part, to their shorter citation windows. The 2026 data should also be interpreted with caution because the search did not cover the entire year.

Citation performance varied across publication years, with older publications generally showing higher average total citation counts than more recent publications (Table 2). The average total citations per document were 109.00 in 1993, 91.15 in 1998, and 61.08 in 2003, compared with 12.34 in 2022, 12.62 in 2023, 5.59 in 2024, 2.53 in 2025, and 0.51 in 2026. Average annual citations also varied across publication years, reaching 4.66 in 2021, compared with 1.26 in 2025 and 0.51 in 2026. These findings should be interpreted cautiously because citation exposure periods differ across publication years. In addition, the 2026 data represent a partial year, as the database search was conducted on 11 September 2026.

Distribution of Publications Across Journals and Bradford’s Law

The 2,704 documents were published across 472 sources, with publications concentrated in a limited number of core journals. According to Bradford’s law, the core zone comprised 9 journals with 940 publications. Journal of Cataract and Refractive Surgery ranked first with 280 publications, followed by Clinical Ophthalmology (139), Ophthalmology (102), British Journal of Ophthalmology (88), and American Journal of Ophthalmology (70). The core zone accounted for 34.8% of all publications. The second and third zones comprised 25 journals with 869 publications and 438 journals with 895 publications, respectively.

Most Productive Authors and Authors’ Citation Performance

Pesudovs K was the most productive author with 53 publications, followed by Lundström M (44), Piñero DP (33), Lamoureux EL (32), and Auffarth GU (29). Kohnen T and Nuijts RMMA each contributed 23 publications, while Hayashi K and Plaza-Puche AB each contributed 20 publications. Alió JL contributed 19 publications (Table 3).

For local citations, Pesudovs K ranked first (935), followed by Lundström M (730), Lamoureux EL (483), Mangione CM (413), Ellwein LB (387), Plaza-Puche AB (350), Stenevi U (346), Bernth-Petersen P (344), Legro MW (324), and Javitt JC (323). Piñero DP ranked 11th with 321 local citations.

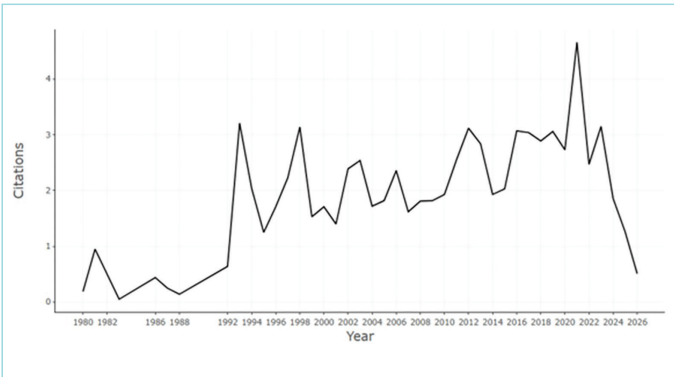


Figure 2. Average citations per year.

Table 2. Citation performance by year			
Year	Number of publications (n)	Average total citations per document	Average annual citations
1993	5	109,00	3,21
1998	27	91,15	3,14
2003	39	61,08	2,54
2012	77	46,73	3,12
2016	122	33,79	3,07
2021	137	27,99	4,66
2022	148	12,34	2,47
2023	157	12,62	3,15
2024	162	5,59	1,86
2025	194	2,53	1,26
2026*	154	0,51	0,51

*The 2026 data are partial and include publications indexed through 11 September 2026.

Table 3. Most productive authors and authors with the highest local citation counts			
A. most productive authors	Number of publications	B. authors with highest local citations	Local citations*
Pesudovs K	53	Pesudovs K	935
Lundström M	44	Lundström M	730
Piñero DP	33	Lamoureux EL	483
Lamoureux EL	32	Mangione CM	413
Auffarth GU	29	Ellwein LB	387
Kohnen T	23	Plaza-Puche AB	350
Nuijts RMMA	23	Stenevi U	346
Hayashi K	20	Bernth-Petersen P	344
Plaza-Puche AB	20	Legro MW	324
Alió JL	19	Javitt JC	323

*Local citations refer to citations received by an author’s publications from other publications within the analyzed dataset.

Scientific Production and International Collaboration by Country

Scientific production was concentrated in the United States (477; 17.64%), China (255; 9.43%), and the United Kingdom (254; 9.39%), followed by Germany (163), Spain (149), and Australia (143). MCP

percentages, calculated relative to the total number of publications within each country, showed that Australia had the highest proportion of internationally co-authored publications (44.06%), followed by Italy (29.33%), Canada (27.47%), and the United Kingdom (25.59%). Türkiye contributed 60 publications (2.22%), including 58 SCP and 2 MCP, corresponding to an MCP proportion of 3.33% (Table 4).

Institutional Scientific Production

The University of London ranked first in institutional productivity with 271 publications, followed by University College London (174), the University of California System (144), Johns Hopkins University (134), and the National University of Singapore (106). Other leading institutions included University of Sydney (102), Moorfields Eye Hospital NHS Foundation Trust (97), Johns Hopkins Medicine (84), University of Toronto (82), and L. V. Prasad Eye Institute (73) (Figure 3).

Keywords and the Temporal Development of Research Themes

Keyword analysis demonstrated a temporal shift in research themes from clinical and functional outcomes toward visual function, quality of life, patient satisfaction, and newer intraocular lens technologies (Figure 4). In the early period, prominent themes included medical outcomes, outcome measures, quality of life, functional impairment, older patients, and visual function. From the mid-2000s onward, visual

function, visual acuity, contrast sensitivity, visual impairment, and psychometric assessment became increasingly prominent. Key terms included contrast sensitivity (302), visual function (356), visual function questionnaire (55), psychometric properties (30), validity (52), and Rasch analysis (51). During the 2010s, research increasingly focused on cataract surgery, visual outcomes, intraocular lens implantation, and patient satisfaction. Among the most frequent terms were cataract-surgery (798), cataract surgery (525), outcomes (453), cataract (397), implantation (273), and patient satisfaction (255). Patient satisfaction became particularly prominent in the later period, which had a median publication year of 2020.

Among the terms concentrated in the more recent publication years (from 2022 onward) were newer intraocular lens technologies and PROMs, including spectacle independence (43), trifocal IOL (25), extended depth (86), extended depth-of-focus (EDOF) (26), EDOF IOL (11), enhanced monofocal IOL (7), and PROMs (5). These terms were interpreted descriptively as recent or emerging topics, based on their temporal distribution within the dataset.

Table 4. Scientific production and international co-authorship by country					
Country	Publications	Publication %	SCP	MCP	MCP (% of country publications)
United States	477	17,64	375	102	21,38
China	255	9,43	213	42	16,47
United Kingdom	254	9,39	189	65	25,59
Germany	163	6,03	125	38	23,31
Spain	149	5,51	120	29	19,46
Australia	143	5,29	80	63	44,06
Japan	123	4,55	119	4	3,25
India	97	3,59	75	22	22,68
Canada	91	3,37	66	25	27,47
Italy	75	2,77	53	22	29,33

MCP (%) was calculated for each country as the number of multiple-country publications divided by the total number of publications, and then multiplied by 100.
SCP: Single-country publications, MCP: Multiple-country publications.

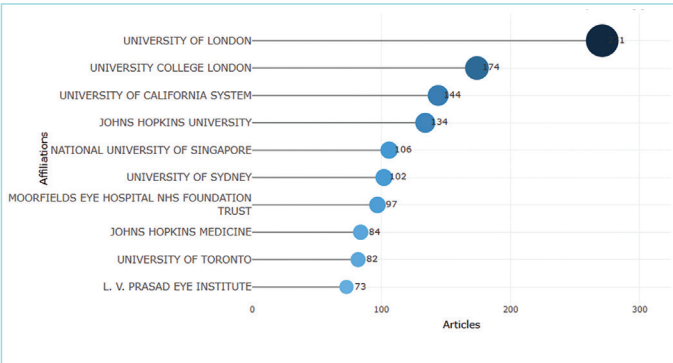


Figure 3. Most relevant affiliations.

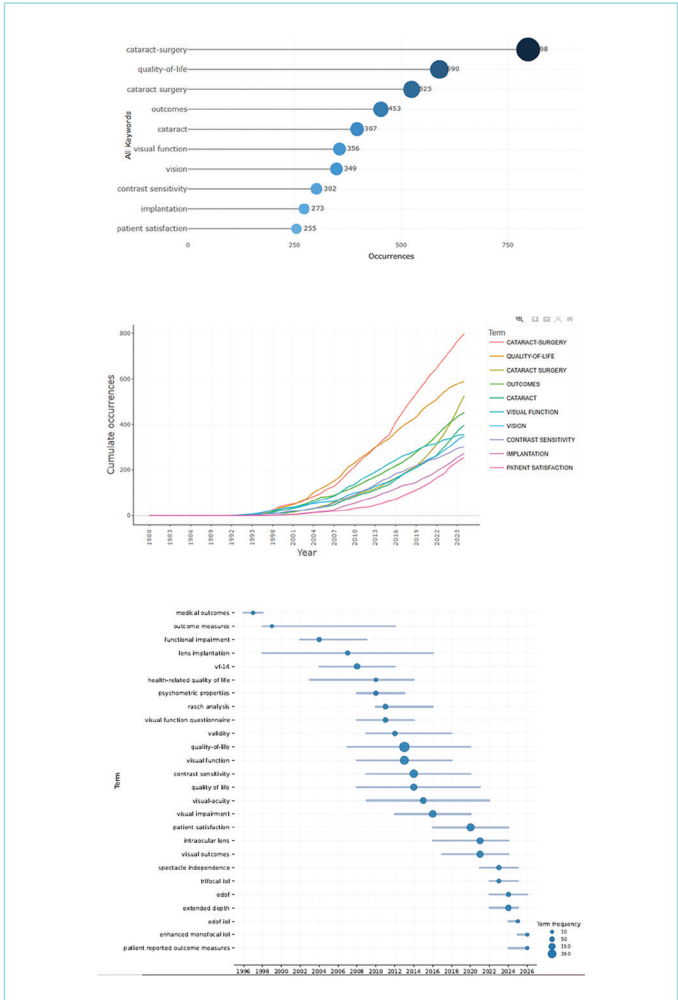


Figure 4. Keyword frequency and temporal evolution of research themes in cataract surgery outcomes. The upper panel shows the most frequent terms, the middle panel illustrates their cumulative frequency over time, and the lower panel presents trend topics according to their temporal distribution.

DISCUSSION

Growth and Expansion of the Research Field

This bibliometric analysis demonstrated sustained growth in research on patient-reported and patient-centered outcomes after cataract surgery, including visual function, quality of life, patient satisfaction, and daily functioning. Scientific production increased markedly from the 2000s onward, reaching 194 publications in 2025. The lower number recorded for 2026 should be interpreted with caution because the database search was conducted on 11 September 2026 and, therefore, does not represent the complete calendar year. The expansion of the field is consistent with a broader approach to cataract surgery outcomes that incorporates patients' experiences alongside conventional clinical measures. Mahmud et al.,¹⁰ in a proposed minimum standard set for cataract surgery outcomes, recommended the inclusion of patient-reported visual function alongside preoperative and postoperative visual acuity. Their framework emphasized that meaningful assessment of cataract surgery should include outcomes that reflect patients' perspectives in addition to clinical measures.

From Objective Visual Outcomes to Patient-Reported Outcomes

The temporal distribution of keywords and trend topics indicated a gradual expansion from visual acuity, visual impairment, and lens implantation toward visual function, quality of life, patient satisfaction, and PROMs. This pattern suggests increasing attention to outcomes that capture patients' perceptions of their visual function and postoperative experience. An important clinical consideration is that visual acuity and patient-perceived visual function do not necessarily represent the same construct. Fung et al.¹¹ found that correlations between PROM scores and visual acuity measures were variable and weak at best. The authors concluded that visual acuity may not fully reflect patients' visual function and suggested that PROMs can complement clinical measures in surgical decision-making and outcome monitoring. These findings provide a clinical context for the bibliometric pattern observed in the present study. The growing prominence of visual function, quality of life, and patient satisfaction suggests that cataract surgery outcomes are increasingly examined from a broader patient perspective rather than through visual acuity alone.

Development of Measurement and Psychometric Approaches

The emergence of terms such as the visual function questionnaire, psychometric properties, validity, and Rasch analysis indicates that research has also focused on how PROs should be measured. The development and validation of the Catquest-9SF represent important examples of this methodological development. Lundström and Pesudovs¹² developed a nine-item Rasch-scaled version of the Catquest questionnaire and demonstrated its validity and responsiveness for measuring visual disability outcomes of cataract surgery. Subsequently, Lundstrom et al.¹³ evaluated the Catquest-9SF before and after cataract surgery and reported excellent psychometric properties, unidimensionality, and high responsiveness to surgical change. These studies illustrate that the development of patient-reported outcome assessment in cataract surgery has involved not only identifying relevant outcomes but also establishing the validity, reliability, and responsiveness of the instruments used to measure them. Thus, the present bibliometric findings suggest a methodological maturation of the field, characterized by increasing attention to standardized and psychometrically sound measurement of patient-reported visual outcomes.

Patient Satisfaction, Spectacle Independence, and New Intraocular Lens Technologies

More recent trends included patient satisfaction, spectacle independence, trifocal intraocular lenses, and EDOF lenses. This shift indicates that research has increasingly addressed outcomes extending beyond postoperative visual acuity, including patients' visual experiences and everyday needs. This pattern is consistent with the systematic review by Wang et al.,¹⁴ which evaluated patient-centered and visual quality outcomes of premium cataract surgery. The review included spectacle independence, dysphotopsia, contrast sensitivity, vision-related quality of life, and patient satisfaction among the outcomes considered. More recent research has continued this approach. For example, Banerjee et al.¹⁵ used PROMs to evaluate patients receiving EDOF intraocular lenses and assessed spectacle use, daily activities, visual symptoms, and satisfaction. The increasing visibility of these topics in this bibliometric analysis therefore reflects a broader research agenda in which the functional and experiential consequences of different intraocular lens technologies are more frequently considered alongside conventional visual measures.

Geographic and Institutional Concentration of Research

Country- and institution-level analyses demonstrated that research production was concentrated in a limited number of countries and research centers. The United States was the most productive country, followed by China and the United Kingdom; the most productive affiliations included the University of London, University College London, the University of California System, and Johns Hopkins University. The overall international co-authorship rate was 22.15%, although substantial variation was observed between countries. This distribution indicates that the research field has developed through a relatively concentrated network of established academic and clinical centers. However, publication volume and international co-authorship should not be interpreted as direct indicators of research quality or clinical impact. The relatively modest overall international co-authorship rate also suggests potential for further multicenter and cross-national collaboration. Institutional findings should be interpreted in light of the structure of WoS affiliation data, as different units within the same research organization may be indexed separately. Nevertheless, the concentration of publications among major universities and specialized ophthalmic centers indicates the presence of established research infrastructures supporting this field.

Overall, the findings demonstrate that research on patient-reported and patient-centered outcomes after cataract surgery has expanded substantially in both volume and scope. The research agenda has progressed from predominantly clinical and visual indicators toward broader assessment of visual function, quality of life, satisfaction, spectacle independence, and other PROs. Future research may benefit from wider sampling.

Study Limitations

This study has several limitations. First, the analysis was based solely on the WoSCore Collection; therefore, relevant publications indexed in other databases may not have been captured. Second, the search strategy and terminology may have excluded studies using alternative terms for cataract surgery or patient-reported and patient-centered outcomes. Third, bibliometric indicators reflect publication and citation patterns but do not directly measure methodological quality, clinical

relevance, or scientific value. Citation-based comparisons should also be interpreted cautiously because older publications have had longer citation windows than recent studies. The 2026 data were collected on 11 September 2026, and therefore do not represent the complete calendar year. Institutional affiliations and keyword patterns may also be influenced by database indexing practices. Accordingly, the findings should be interpreted as a bibliometric representation of the available literature rather than a comprehensive assessment of the quality or clinical impact of the research in the field.

Although the search strategy was designed to comprehensively capture patient-reported and patient-centered outcomes, the use of broad Boolean terms may have resulted in the inclusion of some false-positive records. A random sample of 100 records was independently reviewed by two researchers to assess this potential source of error; however, because the entire dataset was not manually screened at the full-text level, some degree of misclassification cannot be completely excluded.

Furthermore, the analyses were descriptive and exploratory, and no inferential statistical testing was performed. Therefore, the observed temporal patterns and associations should not be interpreted as statistically significant differences or causal relationships.

CONCLUSION

This bibliometric and science-mapping study demonstrated substantial growth in research on patient-reported and patient-centered outcomes after cataract surgery from 1980 to 2026. The research agenda has progressively expanded beyond conventional visual outcomes toward visual function, quality of life, patient satisfaction, spectacle independence, and outcomes associated with newer intraocular lens technologies. The emergence of psychometric evaluation and Rasch analysis also reflects increasing methodological attention to the measurement of PROs. Scientific production was concentrated in a limited number of countries and academic and clinical research centers, while the international co-authorship rate of 22.15% indicates potential for greater cross-national collaboration. Future research should prioritize standardized PROMs, broader international and multicenter collaboration, and the integration of multiple bibliographic databases to provide a more comprehensive understanding of the field.

MAIN POINTS

- The scientific literature on patient-reported and patient-centered outcomes after cataract surgery has shown sustained growth from 1980 to 2026.
- Research has progressively expanded beyond visual acuity toward quality of life, visual function, patient satisfaction, and broader patient-centered outcome domains.
- Psychometric evaluation, Rasch analysis, and standardized patient-reported outcome measures (PROMs) have emerged as important methodological themes.
- Research production is concentrated in a limited number of countries, institutions, and journals, with international co-authorship accounting for 22.15% of publications.
- Future research should promote the standardized use of PROMs and broader international, multicenter collaboration.

ETHICS

Ethics Committee Approval: This study was based exclusively on data retrieved from the Web of Science Core Collection database and did not involve human participants, animals, or patient records. Therefore, ethical approval and informed consent were not required.

Informed Consent: Not applicable. This study was based exclusively on publicly available bibliographic data and did not involve human participants.

Footnotes

Authorship Contributions

Concept: D.K., G.D.S., Design: D.K., G.D.S., Data Collection and/or Processing: D.K., G.D.S., Analysis and/or Interpretation: D.K., G.D.S., Literature Search: D.K., G.D.S., Writing: D.K., G.D.S.

DISCLOSURES

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

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

Declaration on the Use of Artificial Intelligence (AI): Artificial intelligence has been used to assist with “language”.

REFERENCES

1. Vision Loss Expert Group of the Global Burden of Disease Study; GBD 2019 Blindness and Vision Impairment Collaborators. Global estimates on the number of people blind or visually impaired by cataract: a meta-analysis from 2000 to 2020. *Eye (Lond)*. 2024; 38(11): 2156-72.
2. Han X, Zhang J, Liu Z, Tan X, Jin G, He M, et al. Real-world visual outcomes of cataract surgery based on population-based studies: a systematic review. *Br J Ophthalmol*. 2023; 107(8): 1056-65.
3. Grimfors M, Lundström M, Hammar U, Kugelberg M. Patient-reported visual function outcome in cataract surgery: test-retest reliability of the Catquest-9SF questionnaire. *Acta Ophthalmol*. 2020; 98(8): 828-32.
4. Hays RD, MacRae S, Holladay J, Tarver ME, Lum F, Stark W, et al. Development of a patient-reported outcome measure to assess symptoms associated with cataract surgery and intraocular lens implants. *Ophthalmology*. 2023; 130(7): 715-25.
5. Sparrow JM, Grzeda MT, Frost NA, Johnston RL, Liu CSC, Edwards L, et al. Cataract surgery patient-reported outcome measures: a head-to-head comparison of the psychometric performance and patient acceptability of the Cat-PROM5 and Catquest-9SF self-report questionnaires. *Eye (Lond)*. 2018; 32(4): 788-95.
6. Flood J, Koul R, Lundström M, Raju J. Patient-reported outcomes measure after cataract surgery using the Catquest-9SF questionnaire in a U.S. Veterans Health Administration population. *J Cataract Refract Surg*. 2026; 52(6): 530-6.
7. Botta J, Barsam A, Dmitriew A, Zaldivar R, Wiley WF, Windsor S. Factors influencing outcome satisfaction after cataract surgery: patient-reported insights from the RayPro database. *BMC Ophthalmol*. 2024; 24(1): 528.
8. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: an overview and guidelines. *Journal of Business Research*. 2021; 133: 285-96.
9. Du Q, Zhao R, Wan Q, Li S, Li H, Wang D, et al. Protocol for conducting bibliometric analysis in biomedicine and related research using CiteSpace and VOSviewer software. *STAR Protoc*. 2024; 5(3): 103269.

10. Mahmud I, Kelley T, Stowell C, Haripriya A, Boman A, Kossler I, et al. A proposed minimum standard set of outcome measures for cataract surgery. *JAMA Ophthalmol.* 2015; 133(11): 1247-52.
11. Fung SS, Luis J, Hussain B, Bunce C, Hingorani M, Hancox J. Patient-reported outcome measuring tools in cataract surgery: clinical comparison at a tertiary hospital. *J Cataract Refract Surg.* 2016; 42(12): 1759-67.
12. Lundström M, Pesudovs K. Catquest-9SF patient outcomes questionnaire: nine-item short-form Rasch-scaled revision of the Catquest questionnaire. *J Cataract Refract Surg.* 2009; 35(3): 504-13.
13. Lundstrom M, Behndig A, Kugelberg M, Montan P, Stenevi U, Pesudovs K. The outcome of cataract surgery measured with the Catquest-9SF. *Acta Ophthalmol.* 2011; 89(8): 718-23.
14. Wang SY, Stem MS, Oren G, Shtein R, Lichter PR. Patient-centered and visual quality outcomes of premium cataract surgery: a systematic review. *Eur J Ophthalmol.* 2017; 27(4): 387-401.
15. Banerjee S, Kaufmann C, Tinner C, Iselin K, Amstutz CA, Hedinger YM, et al. Patient-reported outcome measures (PROMs) with refractive and diffractive extended depth of focus (EDOF) intraocular lenses. *Klin Monbl Augenheilkd.* 2024; 241(4): 369-73. English.