e-ISSN: 2149-3189

European Research Journal

Volume 12 Issue 7 July 2026

Available Online at https://www.eurj.org.tr

Published by Nicaea Medical Publishing

The European Research Journal

Aim and Scope

The European Research Journal (EuRJ) is an international, independent, double-blind peer reviewed, Open Access and online publishing journal, which aims to publish papers on all the related areas of basic and clinical medicine.

Editorial Board of the European Research Journal complies with the criteria of the International Council of Medical Journal Editors (ICMJE), the World Association of Medical Editors (WAME), and Committee on Publication Ethics (COPE).

The journal publishes a variety of manuscripts including original research, case reports, invited review articles, technical reports, how-to-do it, interesting images and letters to the editor. The European Research Journal has signed the declaration of the Budapest Open Access Initiative. All articles are detected for similarity or plagiarism. Publication language is English.

EuRJ recommends that all of our authors obtain their own ORCID identifier which will be included on their article.

The journal is published (January, February, March, April, May, June, July, August, September, October, November and December).

Abstracting and Indexing

The journal is abstracted and indexed with the following: ULAKBIM TR Index (ULAKBİM TR DİZİN), NLM Catalog (NLM ID: 101685727), Google Scholar (h-index: 15), EMBASE, ProQuest Central, EBSCO Academic Search Ultimate, J-Gate, EZB, TURK MEDLINE, Turkish Citation Index, ResearhGate, SOBIAD, ScienceGate, Publons, (Clarivate Web of Science)

Publisher

The European Research Journal (EuRJ)
Nicaea Medical Publishing
Konak Mh. Kudret Sk. Şenyurt İş Mrk. Blok No:6 İç kapı no: 3
Nilüfer/Bursa-Türkiye
info@nicaeamp.com

Available Online at https://www.eurj.org.tr
https://www.nicaeamp.com

e-ISSN: 2149-3189

EDITORIAL BOARD

EDITOR-IN-CHIEF

EDITORS

OWNER ON BEHALF OF THE PRUSA MEDİCAL PUBLISHING

ASSISTANT EDITOR

SECTION EDITORS

Meliha KASAPOGLU AKSOY, MD.,
Associate Professor,
University of Health Sciences, Bursa Yuksek Ihtisas Training & Research Hospital,
Department of Physical Therapy and Rehabilitation,
Bursa, Türkiye

Arda ISIK, MD.,
Associate Professor,
Medeniyet University School of Medicine,
Department of General Surgery,
Istanbul, Türkiye

Kadir Kaan OZSIN, MD.,
Associate Professor,
University of Health Sciences, Bursa Yuksek Ihtisas Training & Research Hospital,
Department of Cardiovascular Surgery,
Bursa, Türkiye

Cihan AYDIN, MD.,
Associate Professor,
Tekirdağ Namık Kemal University, Faculty of Medicine,
Department of Cardiology,
Tekirdağ, Türkiye

Sayad KOCAHAN, PhD.,
Professor,
University of Health Sciences, Gülhane Medical Faculty,
Department of Physiology,
Ankara, Türkiye

Gokhan OCAKOGLU, PhD.,
Professor,
Uludag University School of Medicine,
Department of Biostatistics,
Bursa, Türkiye

Nurullah DOGAN, MD.,
Professor,
İstanbul Atlas University School of Medicine,
Department of Radiology,
Bursa, Türkiye

Ömer Faruk KARATAS, PhD.,
Professor,
Erzurum Technical University,
Department of Molecular Biology and Genetics,
Erzurum, Türkiye

Serhat YALÇINKAYA, PhD.,
Associate Professor,
Priavate Bursa NEV Health Group,
Department of Thoracic Surgery,
Bursa, Türkiye

Evren DİLEKTAŞLI, MD.,
Professor,
VM Medical Park Bursa Hospital
Department of General Surgery,
Bursa, Türkiye

Nurcan ÖZYAZICIOĞLU,
Professor,
Department Nursing and Health Sciences
Bursa Uludağ Unibersity
Bursa, Türkiye

Burcu DİNÇGEZ, MD.,
Professor,
University of Health Sciences, Bursa Yuksek Ihtisas Training & Research Hospital,
Department of Gynecology and Obstetrics,
Bursa, Türkiye

Yenal DUNDAR, MD.,
Consultant Psychiatrist
University of Liverpool,
Liverpool, UK

Başar CANDER, MD.,
Professor,
Bezmialem Vakif University,
Department of Emergency Medicine
İstanbul, Türkiye

Aylin COLPAN, MD.,
Associate Professor,
Jefferson University-Lehigh Valley Hospital
Department of Infectious Diseases
Allentown, ABD

Sanjiv RAMPAL, MD.,
Associate Professor,
International Medical University
Department of Orthopaedics, Sports Medicine
Kuala Lumpur, Malaysia

Table of Contents

Original Articles

Case Report

The European Research Journal 2026;12(7):708-718 Research Article

DOI: https://doi.org/10.18621/eurj.1817709 *Obstetrics and Gynaecology*

Effects of the Enhanced Recovery After Surgery Protocol on Postoperative Recovery, Sleep Quality, and Postpartum Depression in Elective Cesarean Section Patients: A Single-Center, Prospective Study

Elif Keleş Tayfur[1], Deha Denizhan Keskin[1]

[1]*Deparment of Obstetrics and Gynecology, Ordu University, Faculty of Medicine, Training and Research Hospital, Ordu, Türkiye*

Abstract:

Objective: To assess the effects of the Enhanced Recovery After Surgery–Cesarean Delivery (ERAS-CD) protocol on early postoperative recovery, pain scores, postpartum sleep quality, and postpartum depression in patients undergoing elective cesarean delivery.

Methods: This single-center prospective cohort study involved women with at least one cesarean delivery who underwent planned CD between December 2022 and May 2023. The primary outcome was to reduce hospital stay and morbidity following CD. Secondary outcomes included breastfeeding/formula supplementation status, postoperative pain score, postpartum sleep quality, and predisposition to postpartum depression. The protocol included preoperative oral intake of cherry juice, multimodal analgesia, antiemetic medications, and early ambulation. The number of patients and minimum sample size per group were determined based on power analysis results from previous studies.

Results: A total of 294 patients were included in the study. The ERAS-CD protocol group and the control group were similar in demographic characteristics, body weight, and parity (P>0.05). Postoperative Visual Analog Scale and severe pain scores were lower in all components of the ERAS-CD protocol group (P<0.001). The total score and two subcomponent scores (Sleep duration and Habitual sleep efficiency) of the Pittsburgh Sleep Quality Index were significantly lower in the ERAS-CD group (P=0.005, P=0.004, and P=0.016; respectively). The total score of the Postoperative Recovery Index, as well as the subcomponent scores for bowel symptoms and general symptoms, were significantly lower in the ERAS-CD group (P=0.041, P=0.020, and P=0.045; respectively). Moreover, the Edinburgh Postnatal Depression Scale score and the score indicating predisposition to depression were significantly lower in the ERAS-CD group (P<0.001 for both).

Conclusion: Consistent with the existing literature, the present study demonstrates that the implementation of the ERAS-CD protocol in Cesarean Deliveries has favorable effects on preoperative, intraoperative, and postoperative outcomes. Specifically, the ERAS-CD protocol was shown to reduce postoperative pain, improve postpartum sleep quality, and decrease the predisposition to postpartum depression.

Keywords: ERAS, Postoperative Recovery Index, Postpartum, Depression, Cesarean Delivery

Submitted: November 4, 2025 **Accepted:** December 20, 2025 **Published Online:** December 26, 2025

How to cite this article: Keleş Tayfur E, Keskin DD. Effects of the Enhanced Recovery After Surgery Protocol on Postoperative Recovery, Sleep Quality, and Postpartum Depression in Elective Cesarean Section Patients: A Single-Center, Prospective Study. *Eur Res J.* 2026;12(7):708-718. doi: 10.18621/eurj.1817709

Corresponding author: Elif Keleş Tayfur, MD., Phone: +90 452 225 01 85, E-mail: keleselif@hotmail.com

Available Online at https://www.eurj.org.tr

Cesarean delivery is the most commonly performed abdominal surgery worldwide. Enhanced Recovery After Surgery (ERAS) is a contemporary concept that encompasses multiple interventions designed to enhance recovery. It is a protocol that covers preoperative, intraoperative, and postoperative interventions and, through evidence-based methods, aims to facilitate earlier hospital discharge [1]. The foundations of the ERAS concept were first laid in the 1990s by Prof. Dr. Henrik Kehlet [2]. The ERAS Society has published the Enhanced Recovery After Surgery - Cesarean Delivery (ERAS-CD) guideline for cesarean deliveries, one of the most frequently performed surgical procedures worldwide [3-5].

The primary aim of the ERAS protocol is to reduce hospital stay and morbidity, and to enable patients to return to their daily activities as soon as possible by implementing methods to improve postoperative outcomes and enhance the recovery process. Over the years, ERAS protocol guidelines have been published across numerous surgical subspecialties, including gynecology and gynecologic oncology [6]. The protocol aims to shorten the postoperative care process and enhance recovery by minimizing the metabolic response to surgical stress [1].

The Pittsburgh Sleep Quality Index (PSQI), developed by Buysse *et al*. [7] in 1989, is a 24-item scale for evaluating sleep quality. The Postoperative Recovery Index (PoRI), developed by Butler *et al*. [8] in 2012, evaluates early postoperative recovery. A literature review revealed no previous studies using PoRI in patients undergoing the ERAS-CD protocol. The primary aim of this study was to investigate whether the ERAS-CD protocol accelerates postoperative recovery, improves postpartum sleep quality, and reduces the incidence of postpartum depression compared to the standard perioperative care protocol, and to compare pain intensity using the Visual Analog Scale (VAS), early postoperative recovery quality using the PoRI, postpartum sleep quality using the PSQI, and postpartum depression predisposition using the Edinburgh Postnatal Depression Scale (EPDS).

METHODS

This prospective, randomized, single-center study was conducted at the Department of Obstetrics and Gynecology, Ordu University Faculty of Medicine, following ethics committee approval (No. 2022/263, dated 25/11/2022). Written and signed informed consent was obtained from all participants before the study in the presence of their relatives. The study was conducted in accordance with the principles of the Declaration of Helsinki.

Two hundred and ninety-four patients who applied to the Department of Obstetrics and Gynecology, Ordu University Faculty of Medicine between 25 November 2022 and 15 May 2023, who were scheduled for elective cesarean section, and who met the inclusion criteria and agreed to participate in the study, were included in the study.

Inclusion criteria were: age ≥18, gestational age ≥37 weeks, at least one prior cesarean delivery, no history of psychiatric disorder or use of psychiatric medication, no pregnancy-related comorbidities (such as hypothyroidism, hyperthyroidism, type 1 or type 2 diabetes mellitus, gestational diabetes, chronic hypertension, or gestational hypertensive disorders), no emergency cesarean due to antepartum bleeding, and administration of regional anesthesia.

Patients who met the inclusion criteria and agreed to participate were randomly assigned by drawing either a burgundy or a blue card from a box. Those who drew a burgundy card were assigned to the ERAS-CD protocol group, while those who drew a blue card were assigned to the control group receiving standard perioperative care. A total of 147 patients were randomized to the ERAS-CD group, and 147 to the control group. The ERAS-CD protocol was explained to the patients scheduled to undergo it in the preoperative, intraoperative, and postoperative periods. Patients in the ERAS-CD group were given 200 mL of pulp-free cherry juice 2 hours before surgery. In addition, patients were allowed to consume clear liquids such as tea and coffee up to 2 h before surgery and solid foods up to 6 h before surgery; their nutrition was continued with carbohydrate-rich fluids. Preoperatively, antacids (Esmoprazole®) and H2 receptor antagonists (Ranitidine®) were administered together. Thirty minutes before skin incision, patients were administered 2 g of first-generation cephalosporin parenterally. No bowel cleansing was performed before the surgery. Patients were fitted with pneumatic compression stockings preoperatively. After spinal

anesthesia was administered in the operating room, urinary catheters were inserted. Spinal anesthesia was administered to the patients. Perioperative fluid therapy was provided at 0.5 ml/kg/min. Intravenous (i.v.) balanced fluid (2,000 mL) containing Ephedrine® was administered. A combination of Zofer® (ondansetron) and Metpamid® (metchloropamide) was given to prevent intraoperative nausea and vomiting. 30 IU Synpitan® (oxytocin) was slowly infused in 1 liter i.v. fluid intraoperatively. For postoperative analgesic prophylaxis, a multimodal analgesic regimen was routinely administered for two days, including Dicloron® (diclofenac sodium, 1 vial=75 mg) 3 times a day intramuscularly and Parol® (paracetamol, 1 vial=500 mg) 3×1 intravenously. Perioperative blood glucose was monitored four times daily. After surgery, the patient was instructed to chew xylitol-containing gum upon arrival in the room. Postoperatively, room temperature was maintained above 23°C, and the patient's body temperature above 36°C. Oral feeding was initiated at the 4th postoperative hour, the urinary catheter was removed, and the patient was mobilized.

In the control group, oral intake was restricted from 00:00 the night before surgery. Preoperative bowel cleansing was not performed. 2 g of first-generation cephalosporin was administered 30-60 min before the surgery, and 1 g of azithromycin was added for patients with water-ruptured membranes. Urinary catheterization was performed before patients were taken to the operating room, and spinal anesthesia was administered to all patients. Conventional perioperative intravenous fluid therapy was provided, and short-acting opioid analgesics (Fentanyl®) were used. 30 IU Sinpitan® was administered intraoperatively in 500 ml i.v. fluid. Postoperatively, patients received Dicloron® intramuscularly and Parol® parenterally. Oral intake was initiated at the 8th postoperative hour, the urinary catheter was removed, and out-of-bed mobilization was performed. These patients were discharged at the 48th postoperative hour after confirming the passage of flatus.

Data Collection

For all patients included in the study, preoperative, postoperative 6th hour, and 24th hour hemogram parameters, albumin, glucose, and C-reactive protein (CRP) values were recorded on the data collection form. Intraoperative nausea and vomiting were assessed and documented. Neonates' weights and APGAR scores at 1 and 5 min were recorded. Postoperative breastfeeding status and the need for formula feeding were assessed and documented. Patients were asked to note the time to pass flatus and report it. Patients were asked when they felt ready for discharge, and this was recorded. After discharge, the duration of urinary catheter placement was obtained from patient records, and the total catheterization time was calculated. Preoperative fasting duration, operation time, and time to oral intake were recorded, and total fasting duration was calculated. Intraoperative and postoperative fluid volumes, operation times, and all anesthetic, analgesic, and antiemetic medications administered with their doses were documented. Patients' average preoperative, perioperative, and postoperative blood pressures were calculated and recorded.

Postoperative pain scores of patients receiving the ERAS-CD protocol and standard perioperative care were recorded at 6, 12, and 24 hours using the internationally accepted Visual Analog Scale (VAS). All patients were contacted on the 10th postoperative day, and the EPDS was administered to assess predisposition to postpartum depression. Scoring was performed according to patient responses. The guidelines prepared by the Public Health Institution of the Ministry of Health of the Republic of Turkey state that the mother's psychological state can be assessed between the first day and the forty-two days after delivery. In this study, we decided that it would be appropriate to evaluate the EPDS score of the patients when they came for their follow-up visit on the 10th day after delivery.

All patients were contacted on the 10th postoperative day, and their postpartum sleep quality was assessed using the PSQI; responses were scored according to PSQI guidelines, and total PSQI scores were calculated. Patients' early postoperative recovery quality was assessed using the PoRI on postoperative day 1. PoRI scores were calculated according to their responses. The PoRI that was tested for validity and reliability by Butler *et al*. [9] in 2012 consists of 37 items. High scores received from the index reflect that more problems are experienced in postoperative recovery, whereas low scores show that postoperative recovery has been easier.

Statistical Analysis

Sample size was calculated using G*Power 3.1.9.6 to achieve 90% power at a 95% confidence level to detect a 20% reduction in primary outcome opioid analgesic use, resulting in a minimum of 108 patients per group, totaling 216 patients.

Statistical analysis was performed using Microsoft® Excel version 21.02 and SPSS version 23.0.0 (SPS Inc.; Chicago, IL, USA). Descriptive data were presented as mean±standard deviation. Normality of continuous variables was assessed using the Kolmogorov-Smirnov and Shapiro-Wilk tests. For comparison of numerical variables, Student's t-test was used for normally distributed data, and the Mann-Whitney U test for non-normally distributed data. Categorical variables were analyzed using Pearson's chi-square test or Fisher's Exact test. A P-value <0.05 was considered statistically significant.

RESULTS

The mean age of patients in the ERAS-CD protocol group was 28.87±5.00 years, and the mean age of the patients in the control group was 29.19±5.16 years. Demographic characteristics of the ERAS-CD protocol group and the control group were similar (P=0.590). The obstetric characteristics of the patients in the ERAS-CD and control groups and the characteristics of the newborns were compared. When groups were compared in terms of the mean number of gravida (2.69±1.06 vs 2.78±1.14; P=0.492), the mean number of parity (2.31±0.5 vs 2.43±0.7; P=0.107), the mean number of caesarean section (2.28±0.5 vs 2.41±0.6; P=0.073), the mean gestational week (38.08±0.7 vs 38.1±0.7; P=0.280), newborn birth weight (3261.5±440.9 vs 3315.1±4 21.5; P=0.287), APGAR 1st minute score (8.78±0.4 vs 8.81±0.4; P=0.571) and APGAR 5th minute score (9.68±0.4 vs 9.76±0.4; P=0.216), no statistically significant difference was found between the control group and the ERAS-CD group.

While the mean preoperative glucose value was 108.27±3 mg/dL in the ERAS-CD group, it was 82.9±16.9 mg/dL in the control group. The difference between the groups was statistically significant (P<0.001).

In the control group, the amount of intraoperative fluid administered was higher than in the ERAS-CD group, and the difference was statistically significant (2198.6±450.4 vs 2000.00±0.000; P<0.001). The amount of the short-acting opioid Fentanyl® used intraoperatively was 8.95±16.4 mcg in the ERAS-CD protocol group and 31.4±19.7 mcg in the control group; the difference was found to be statistically significant (P<0.001). The mean intraoperative systolic blood pressure was 107.4±10.1 mmHg in the control group and 111.1±8.5 mmHg in the ERAS-CD group, showing a statistically significant difference

TABLE 1. Comparison of Intraoperative Data of ERAS-CD and Control Groups

	ERAS-CD group (n=147)	Control group (n=147)	P-value[a]
Liquid amount (mL)	2000.00±0.000	2198.64±450.416	**<0.001**
Fentanyl® amount (mcg)	8.95±16.451	31.46±19.761	**<0.001**
Dormicum® amount (mg)	1.00±0.00	0.46±0.500	**<0.001**
Propofol® (mg)	14.29±36.882	20.54±49.911	0.222
Synpitan® amount (IU)	31.29±3.886	31.63± 5.141	0.523
Ephedrine® amount (mg)	24.93±5.703	7.14±11.302	**<0.001**
Systolic blood pressure (mmHg)	111.10±8.539	107.49±10.115	**0.001**
Diastolic blood pressure (mmHg)	66.14±8.654	62.21±9.715	**<0.001**
Operation time (min)	37.21±11.167	39.13±9.153	0.108

Data are shown as mean±standard deviation. ERAS-CD, enhanced recovery after surgery–cesarean delivery.
[a]Student's t test. Statistically significant P-values are shown in bold.

TABLE 2. Comparison of Postoperative 6th Hour and Postoperative 24th Hour Biochemical Parameters and Postoperative Data

	ERAS-CD group (n=147)	Control group (n=147)	P-value[a]
Glucose 6th hour(mg/dL)	99.18±24.258	79.54±14.659	**<0.001**
Albumin 24th hour (g/dL)	30.382±2.9053	29.592±3.3251	**0.031**
Glucose 24th hour (mg/dL)	92.90±18.547	85.35±14.260	**<0.001**
CRP 24th hour(mg/L)	51.713±46.9935	56.508±37.1092	0.332
Liquid amount (mL)	3,000±0.000	2996.60±41.239	0.318[a]
Dichloron® amount (flacon)	6.00±0.000	2.43±1.250	**<0.001**
Parol® amount (flacon)	6.00±0.000	0.43±0.876	**0.001**
Aldolan® amount (facon)	0.00±0.00	0.02±0.142	0.082
Systolic blood pressure (mmHg)	111.71±7.239	109.65±9.004	**0.031**
Diastolic blood pressure (mmHg)	67.35±6.672	65.89±6.711	0.063
Total catheter urinary drainage time (min)	240.00±0.000	564.12±66.761	**<0.001**
Total fasting time (min)	360.00 ±0.000	1186.12±315.392	**<0.001**
Gas discharge time(h)	11.20±5.623	14.63±10.878	**<0.001**
Time to feel ready for discharge (h)	21.16±5.024	46.72±7.049	**<0.001**

Data are shown as mean±standard deviation. ERAS-CD, enhanced recovery after surgery–cesarean delivery.
[a]Student's t test. Statistically significant P-values are shown in bold.

(P<0.001). The amount of Dormicum® used intraoperatively was significantly higher in the ERAS-CD group compared to the control group (1.00±0.00 vs 0.46±0.5; P<0.001). Patients in the ERAS-CD group received more Dormicum® and Ephedrine®, while patients in the control group received more Fentanyl® and a higher volume of intravenous fluids. Blood pressure values also indicated that both systolic and diastolic pressures were higher in the ERAS-CD group. There was no significant difference between the groups in terms of Propofol® use, and the requirements for® Synpitan® were met. The implementation of the ERAS-CD protocol did not affect operative time.

The incidence of postoperative complaints was lower in the ERAS-CD group, and this difference was statistically significant.

The ERAS-CD group and the control group were compared in terms of the postoperative 6th-hour hemogram parameters. When the ERAS-CD group

TABLE 3. Comparison of Postoperative 0th Hour, 6th Hour and 24th Hour VAS Scores and EPDS Scores

	ERAS-CD group (n=147)	Control group (n=147)	P-value[a]
VAS 0th hour	3.52±2.925	5.43±3.125	**<0.001**
VAS 6th hour	3.60±2.083	5.56±2.559	**<0.001**
VAS 24th hour	1.73±1.698	3.76±2.525	**<0.001**
EPDS score	5.58±4.420	14.14±4.453	**<0.001**

Data are shown as mean±standard deviation. ERAS-CD, enhanced recovery after surgery–cesarean delivery; VAS, visual analog scale; EPDS, Edinburgh postpartum depression scale.
[a]Student's t test. Statistically significant P-values are shown in bold.

and the control group were compared, no statistical difference was found between the groups. No significant difference was observed between the groups in terms of postoperative 24th hour hemogram parameters.

Postoperative 6th-hour glucose and postoperative 24th-hour albumin, glucose, and CRP values of ERAS-CD and control groups were compared. The postoperative 24th hour albumin level was 30.3±2.9 in the ERAS-CD group, while it was 29.5±3.3 g/ dl in the control group, showing a statistically significant difference (P=0.031. Both 6th- and 24th-hour glucose levels were higher in the ERAS-CD group compared to the control group.

The amounts of Dicloron® and Parol® used postoperatively, and the mean postoperative systolic blood pressure differed significantly between the ERAS-CD and control groups (Table 1).

In accordance with the ERAS-CD protocol, patients in this group received higher doses of Dicloron® and Parol®. Mean postoperative systolic and diastolic blood pressures were higher in the ERAS-CD group. The difference in systolic blood pressure was statistically significant.

The time to first postoperative flatus (11.2±5.6 vs. 14.6±10.8; P=0.001), total fasting duration (360.0±0.0 vs. 1186.1±315.3; P<0.001), total urinary catheter drainage duration (240.0±0.0vs. 564.1±66.7; P<0.001), and the time to feel ready for discharge (21.1±5.0 vs. 46.7±7.0; P<0.001), were all statistically significantly different between the ERAS-CD and the control group (Table 2).

In summary, total urinary catheter drainage time and total fasting duration were significantly shorter in the ERAS-CD group. The time to first flatus and the time patients felt ready for discharge were also shorter in the ERAS-CD group compared to the control group. In the ERAS-CD group, 93.9% (n=138) of patients were breastfeeding, compared to 85.7% (n=126) in the control group, and the difference was statistically significant (P=0.021). Regarding formula supplementation, 37.4% (n=55) of infants in the control group required formula supplementation, compared to 21.8% (n=32) in the ERAS-CD group. The difference between the groups was also statistically significant (P=0.003). Breastfeeding rates were higher and formula supplementation lower in the ERAS-CD group.

Postoperative 0-hour, 6-hour, and 24-hour VAS scores of the patients in the ERAS-CD group and the control group were compared. Patients in the ERAS-CD group experienced less pain.

When the patient groups were compared in terms of postpartum depression tendencies, the mean EPDS score was 5.5±4.4 in the ERAS-CD group and 14.14±4.4 in the control group, and the difference was statistically significant (P<0.001) (Table 3).

According to these results, it was determined that the patients in the ERAS-CD protocol group were less prone to postpartum depression than the control group. An increase in predisposition to postpartum depression was observed in 12 patients (8.2%) in the ERAS-CD group, compared to 95 patients (64.6%) in the control group, with the difference being statistically

TABLE 4. Comparison of PSQI Scoring and Subcomponents

	ERAS-CD group (n=147)	Control group (n=147)	P-value[a]
Total PSQI	9.56±2.809	10.52±2.922	**0.005**
Subjective sleep quality	1.84±0.641	1.93±0.746	0.242
Sleep latency	1.12±0.711	1.20±0.728	0.373
Sleep duration	1.52±0.797	1.78±0.745	**0.004**
Habitual sleep efficiency	1.37±1.212	1.72±1.232	**0.016**
Sleep disorder	2.05±0.783	2.18±0.777	0.156
Daytime dysfunction	1.65±0.782	1.70±0.823	0.611

Data are shown as mean±standard deviation. ERAS-CD, enhanced recovery after surgery–cesarean delivery; PSQI, Pittsburgh sleep quality index. [a]Student's t test. Statistically significant P-values are shown in bold.

TABLE 5. Comparison of the Scores of PoRI and Its Sub-dimensions

	ERAS-CD group (n=147)	Control group (n=147)	P-value[a]
PoRI total	2.06660±0.547213	2.20973±0.644434	**0.041**
Psychological symptoms	2.0306±0.59744	2.1190±0.71812	0.252
Physical activities	2.86565±0.724907	2.95408±0.847675	0.337
Appetite symptoms	1.6122±0.68633	1.7126±0.69757	0.215
Bowel symptoms	2.137±0.9114	2.404±1.0372	**0.020**
General symptoms	1.6871±0.73200	1.8588±0.72920	**0.045**

Data are shown as mean±standard deviation. ERAS-CD, enhanced recovery after surgery–cesarean delivery; PoRI, postoperative recovery index.

[a]Student's t test. Statistically significant P-values are shown in bold.

significant (P<0.001).

PSQI scores and sub-components were compared between the ERAS-CD and control groups. When the patient groups were compared in terms of postoperative sleep quality, PSQI score was 9.56±2.809 in the ERAS-CD group, while it was 10.5±2.9 in the control group, and the difference between the groups was statistically significant (P=0.005). It is seen that the patients in the ERAS-CD group achieved better sleep quality (Table 4).

When the groups were compared about postoperative recovery, the difference in total PoRI score (2.06±0.5 vs. 2.2±0.6; p=0.041) was found to be statistically significant. It was concluded that the patients in the ERAS-CD group felt better postoperatively (Table 5).

A comparison was made between the ERAS-CD and control groups regarding the difficulty levels of the PoRI psychological symptoms subdimension. In the ERAS-CD group, 1 patient (0.7%) experienced extreme difficulty, compared to 8 patients (5.4%) in the control group (P=0.042) (Table 6).

DISCUSSION

The rising prevalence of cesarean sections further emphasized the need to improve perioperative care processes during cesarean procedures. The ERAS-CD protocol aims to reduce stress, enhance postoperative responses, and rapidly restore metabolic functions [9]. In a Cochrane review including 27 studies, preoperative carbohydrate fluid loading as an element of the ERAS protocol was shown to shorten the time to return of bowel function, time to first flatus, and length of hospital stay [10]. In a study by Orji *et al*. [11]

TABLE 6. PoRI - Psychological Symptoms of ERAS-CD and Control Groups

		ERAS-CD group		Control group		Total		P-value
		n	%	n	%	n	%	
PoRI - Psychological symptoms	No difficulty	14	9.5	19	12.9	33	11.2	**0.042[b]**
	Little difficulty	8	5.4	4	2.7	12	4.1	
	Medium difficulty	87	59.2	71	48.3	158	53.7	
	Considenable difficulty	37	25.2	45	30.6	82	27.9	
	Extreme difficulty	1	0.7	8	5.4	9	3.1	

ERAS-CD, enhanced recovery after surgery–cesarean delivery; PoRI, postoperative recovery index.

[b]Pearson Chi-Square test. Statistically significant P-values are shown in bold.

comparing early feeding and late feeding in cesarean section patients, time to first flatus and length of hospital stay were found to be shorter in the early feeding group. In a 2013 meta-analysis by Hsu *et al*. [12], including 17 studies with a total of 2,966 cesarean section patients, early postoperative oral feeding was shown to reduce the time to first flatus. In our study, encouraging patients in the ERAS-CD group to consume clear fluids up to 2 hours before surgery and initiating oral feeding 4 hours postoperatively resulted in a shorter total fasting time in this group. Consistent with the literature, time to first flatus was also shorter in the ERAS-CD group compared to the control group.

Several studies in the literature have similarly demonstrated that patients with shorter preoperative fasting times also had shorter hospital stays [13-15]. The ERAS-CD protocol recommends controlled perioperative fluid therapy. In the study conducted by Miller *et al*., the control group received intravenous fluid therapy compared to patients who underwent the ERAS protocol. Aggressive fluid therapy was shown to delay the return of bowel function and increase postoperative ileus, postoperative nausea, vomiting, and length of hospital stay [16]. Our study supports the literature. Patients in the ERAS-CD group received controlled perioperative fluids, and the time to first flatus was observed to be shorter compared to the control group.

Chewing gum was shown to reduce postoperative time to first flatus, accelerate bowel movements, and shorten hospital stay [17, 18]. Various studies on the ERAS protocol have also emphasized reduced hospital stay and complications with its implementation [19-21]. In our study, in accordance with the literature, patients chewing gum and receiving the ERAS-CD protocol had a statistically significantly shorter time to first flatus and time to feeling ready for discharge.

Examining studies on the postoperative ERAS-CD protocol, early urinary catheter removal and mobilization are recommended; a meta-analysis by Huang *et al*. [22] of eight gynecological surgery studies showed that catheter removal within 6 hours reduced urinary tract infections and urinary retention. In our study, urinary drainage duration was shorter; none of the patients in either the ERAS-CD group, which received early mobilization, or the standard perioperative care group required re-catheterization. Our study demonstrated that early removal of the urinary catheter and early mobilization are safe practices. In our study, the ERAS-CD protocol group showed a shorter postoperative time to first flatus, which was associated with significantly lower pain scores, higher patient satisfaction, increased breastfeeding rates, and a reduced time to readiness for discharge.

Pain is the most common and significant complaint after a cesarean section. Effective intraoperative and postoperative analgesia facilitates postoperative recovery and the return of functions. The ERAS-CD protocol recommends opioid-sparing multimodal analgesic prophylaxis. In the relevant literature, Kleiman *et al*. [23] examined the superiority of the ERAS protocol implementation over the control group in cesarean deliveries and demonstrated reductions in opioid consumption, pain scores, and length of hospital stay. A study reported that patients who underwent cesarean delivery and received the ERAS protocol experienced reduced postoperative pain, shorter hospital stays, and lower opioid requirements, while their newborns had better outcomes compared to those in the conventional care control group [24]. In our study, patients in the ERAS-CD protocol group showed statistically significant reductions in VAS scores, the incidence of severe pain, and the time to feeling ready for discharge. Additionally, breastfeeding rates were higher, and formula supplementation rates were lower in this group.

Studies have also reported that VAS scores were significantly lower in the ERAS protocol group [25, 26]. In our study, patients in the standard perioperative care group experienced more severe pain compared to those in the ERAS-CD protocol group, and the difference between the groups was statistically significant.

Apart from our study, no other ERAS-CD protocol study in the literature has been found to utilize both the PSQI and the PoRI. In the study conducted by Cengiz *et al*. [27], high scores received from the index reflect that more problems are experienced in postoperative recovery, whereas low scores show that postoperative recovery has been easier. We also based our study on these criteria. Our study demonstrated that implementation of the ERAS-CD protocol in cesarean patients improved early postoperative

outcomes. When comparing patients in the ERAS-CD protocol group with those in the control group in terms of total PoRI scores, the difference between the groups was found to be statistically significant. Analysis of PoRI subscales revealed significant differences between the groups in the bowel symptoms and general symptoms subscales.

It has been stated that the ERAS-CD protocol implementation may reduce postpartum depression in women [28]. In our study, it was shown that sleep quality was better in the ERAS-CD protocol group compared to the standard perioperative care group. The total EPDS score was also lower in the ERAS-CD protocol group, with fewer patients showing a predisposition to depression.

Strengths and Limitations

In this study, we aimed to provide a previously unexamined contribution to the literature by examining the ERAS protocol in post-cesarean section patients using multiple parameters. We believe our study makes a significant contribution to the literature because it utilized several internationally accepted scoring systems. Apart from our study, no other ERAS-CD protocol study in the literature uses the Pittsburgh Sleep Quality Index and Postoperative Recovery Indices.

There are several limitations that should be considered when interpreting the results of this study. The single - center design may limit the generalizability of the findings to different healthcare settings or patient populations. In addition, postpartum depression and sleep quality were assessed at a single time point in the early postpartum period; therefore, long-term psychological outcomes could not be evaluated.

CONCLUSION

Beyond physical recovery, this protocol improved postpartum sleep quality and significantly reduced predisposition to postpartum depression, while also supporting favorable neonatal outcomes such as higher breastfeeding rates. These findings highlight the ERAS-CD protocol as an effective, patient-centered perioperative care model that enhances both maternal recovery and psychological well-being following cesarean delivery. In conclusion, the implementation of the ERAS-CD protocol in elective cesarean deliveries significantly improved postoperative recovery.

Ethics Approval and Consent to Participate

This study was approved by the Ordu University Clinical Research Ethics Committee (Decision No: 2022/23-263-; date: 25.11.2022). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Written and signed informed consent was obtained from all participants before the study in the presence of their relatives.

Data Availability

All data generated or analyzed during this study are included in this published article. The data that support the findings of this study are available on request from the corresponding author, upon reasonable request.

Authors' Contribution

Study Conception: DDK; Study Design: DDK; Supervision: EKT; Funding: EKT; Materials: EKT; Data Collection and/or Processing: EKT; Statistical Analysis and/or Data Interpretation: DDK; Literature Review: EKT; Manuscript Preparation: EKT; and Critical Review: DDK.

Conflict of Interest

The author(s) disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The author(s) disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgments

The authors have no acknowledgments to declare.

Generative Artificial Intelligence Statement

The author(s) declare that no artificial intelligence-based tools or applications were used

during the preparation process of this manuscript. The all content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

Editor's Note

All statements made in this article are solely those of the authors and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Kehlet H, Wilmore DW. Evidence-based surgical care and the evolution of fast-track surgery. Ann Surg. 2008;248(2):189-198. doi: 10.1097/SLA.0b013e31817f2c1a.
2. Kehlet H. Multimodal approach to control postoperative pathophysiology and rehabilitation. Br J Anaesth. 1997;78(5):606-617. doi: 10.1093/bja/78.5.606.
3. Caughey AB, Wood SL, Macones GA, et al. Guidelines for intraoperative care in cesarean delivery: Enhanced Recovery After Surgery Society Recommendations (Part 2). Am J Obstet Gynecol. 2018;219(6):533-544. doi: 10.1016/j.ajog.2018.08.006.
4. Macones GA, Caughey AB, Wood SL, et al. Guidelines for postoperative care in cesarean delivery: Enhanced Recovery After Surgery (ERAS) Society recommendations (part 3). Am J Obstet Gynecol. 2019;221(3):247.e1-247.e9. doi: 10.1016/j.ajog.2019.04.012.
5. Wilson RD, Caughey AB, Wood SL, et al. Guidelines for Antenatal and Preoperative care in Cesarean Delivery: Enhanced Recovery After Surgery Society Recommendations (Part 1). Am J Obstet Gynecol. 2018;219(6):523.e1-523.e15. doi: 10.1016/j.ajog.2018.09.015.
6. Thorell A, MacCormick AD, Awad S, et al. Guidelines for Perioperative Care in Bariatric Surgery: Enhanced Recovery After Surgery (ERAS) Society Recommendations. World J Surg. 2016;40(9):2065-2083. doi: 10.1007/s00268-016-3492-3.
7. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193-213. doi: 10.1016/0165-1781(89)90047-4.
8. Butler SF, Black RA, Techner L, Fernandez K C, Brooks D, Wood M. Development and validation of the post-operative recovery index for measuring quality of recovery after surgery. J Anesth Clin Res. 2012;3(12):1-8. doi: 10.4172/2155-6148.1000267.
9. Nelson G, Bakkum-Gamez J, Kalogera E, et al. Guidelines for perioperative care in gynecologic/oncology: Enhanced Recovery After Surgery (ERAS) Society recommendations-2019 update. Int J Gynecol Cancer. 2019;29(4):651-668. doi: 10.1136/ijgc-2019-000356.
10. Smith MD, McCall J, Plank L, Herbison GP, Soop M, Nygren J. Preoperative carbohydrate treatment for enhancing recovery after elective surgery. Cochrane Database Syst Rev. 2014;2014(8):CD009161. doi: 10.1002/14651858.CD009161.pub2.
11. Orji EO, Olabode TO, Kuti O, Ogunniyi SO. A randomised controlled trial of early initiation of oral feeding after cesarean section. J Matern Fetal Neonatal Med. 2009;22(1):65-71. doi: 10.1080/14767050802430826.
12. Hsu YY, Hung HY, Chang SC, Chang YJ. Early oral intake and gastrointestinal function after cesarean delivery: a systematic review and meta-analysis. Obstet Gynecol. 2013;121(6):1327-1334. doi: 10.1097/AOG.0b013e318293698c.
13. Awad S, Varadhan KK, Ljungqvist O, Lobo DN. A meta-analysis of randomised controlled trials on preoperative oral carbohydrate treatment in elective surgery. Clin Nutr. 2013;32(1):34-44. doi: 10.1016/j.clnu.2012.10.011.
14. Campos SBG, Barros-Neto JA, Guedes GDS, Moura FA. Pre-Operative Fasting: Why Abbreviate. Arq Bras Cir Dig. 2018;31(2):e1377. doi: 10.1590/0102-672020180001e1377.
15. Pinto Ados S, Grigoletti SS, Marcadenti A. Fasting abbreviation among patients submitted to oncologic surgery: systematic review. Arq Bras Cir Dig. 2015;28(1):70-73. doi: 10.1590/S0102-67202015000100018.
16. Miller TE, Thacker JK, White WD, et al. Reduced length of hospital stay in colorectal surgery after implementation of an enhanced recovery protocol. Anesth Analg. 2014;118(5):1052-1061. doi: 10.1213/ANE.0000000000000206.
17. Short V, Herbert G, Perry R, et al. Chewing gum for postoperative recovery of gastrointestinal function. Cochrane Database Syst Rev. 2015;2015(2):CD006506. doi: 10.1002/14651858.CD006506.pub3.
18. Park SH, Choi MS. Meta-Analysis of the Effect of Gum Chewing After Gynecologic Surgery. J Obstet Gynecol Neonatal Nurs. 2018;47(3):362-370. doi: 10.1016/j.jogn.2018.01.011.
19. Tanaka R, Lee SW, Kawai M, et al. Protocol for enhanced recovery after surgery improves short-term outcomes for patients with gastric cancer: a randomized clinical trial. Gastric Cancer. 2017;20(5):861-871. doi: 10.1007/s10120-016-0686-1.
20. Teixeira UF, Fontes PRO, Conceição CWN, et al. Implementation of Enhanced Recovery After Colorectal Surgery (ERAS) Protocol: Initial Results of The First Brazilian Experience. Arq Bras Cir Dig. 2019;32(1):e1419. doi: 10.1590/0102-672020180001e1419.
21. Desiderio J, Stewart CL, Sun V, et al. Enhanced Recovery after Surgery for Gastric Cancer Patients Improves Clinical Outcomes at a US Cancer Center. J Gastric Cancer. 2018;18(3):230-241. doi: 10.5230/jgc.2018.18.e24.
22. Huang H, Dong L, Gu L. The timing of urinary catheter removal after gynecologic surgery: A meta-analysis of randomized controlled trials. Medicine (Baltimore). 2020;99(2):e18710. doi: 10.1097/MD.0000000000018710.
23. Kleiman AM, Chisholm CA, Dixon AJ, et al. Evaluation of the impact of enhanced recovery after surgery protocol implementation on maternal outcomes following elective cesarean delivery. Int J Obstet Anesth. 2020;43:39-46. doi: 10.1016/j.ijoa.2019.08.004.
24. Patel K, Zakowski M. Enhanced Recovery After Cesarean:

Current and Emerging Trends. Curr Anesthesiol Rep. 2021;11(2):136-144. doi: 10.1007/s40140-021-00442-9.
25. Peng J, Dong R, Jiao J, et al. Enhanced Recovery After Surgery Impact on the Systemic Inflammatory Response of Patients Following Gynecological Oncology Surgery: A Prospective Randomized Study. Cancer Manag Res. 2021;13:4383-4392. doi: 10.2147/CMAR.S294718.
26. Kim MK, Kim JG, Lee G, et al. Comparison of the effects of an ERAS program and a single-port laparoscopic surgery on postoperative outcomes of colon cancer patients. Sci Rep. 2019;9(1):11998. doi: 10.1038/s41598-019-48526-1.
27. Cengiz H, Aygin D. Validity and reliability study of the Turkish version of the Postoperative Recovery Index of patients undergoing surgical intervention. Turk J Med Sci. 2019;49(2):566-573. doi: 10.3906/sag-1806-33.
28. Bollag L, Lim G, Sultan P, et al. Society for Obstetric Anesthesia and Perinatology: Consensus Statement and Recommendations for Enhanced Recovery After Cesarean. Anesth Analg. 2021;132(5):1362-1377. doi: 10.1213/ANE.0000000000005257.

The European Research Journal 2026;12(7):719-729 Research Article

DOI: https://doi.org/10.18621/eurj.1843405 *Radiology and Organ Imaging*

The Effect of Digital Breast Tomosynthesis on BI-RADS Categorization in Different Breast Densities: A Retrospective Evaluation

Deniz Esin Tekcan Şanlı[1], Emre Aksu[1], Ahmet Necati Şanlı[2], Bilal Turan[3]

[1]Department of Radiology, Faculty of Medicine, Gaziantep University, Gaziantep, Türkiye; [2]Department of General Surgery, Abdulkadir Yüksel State Hospital, Gaziantep, Türkiye; [3]Department of General Surgery, Faculty of Medicine, Süleyman Demirel University, Isparta, Türkiye

Abstract:

Objective: This study aimed to evaluate the impact of digital breast tomosynthesis (DBT) on BI-RADS categorization compared to conventional mammography (MMG), across different mammographic breast density types.

Methods: In this retrospective study, 520 female patients aged 35–85 years who underwent both MMG and DBT between 2023 and 2025 were included. Standard craniocaudal (CC) views were acquired for MMG and mediolateral oblique (MLO) views for DBT. Synthetic MLO (sMLO) images were automatically generated by the device from DBT data. Patients were categorized according to the breast density (Type A–D), and all images were reviewed in consensus by two experienced radiologists. Breast Imaging Reporting and Data System (BI-RADS) categories, lesion types, were recorded. BI-RADS categorizations from MMG and DBT were compared overall and within density subgroups.

Results: The mean age was 53.0 ± 9.9 years. Breast density distribution was: Type A (3.8%, n=20), B (28.5%, n= 148), C (43.5%, n=226), and D (24.2%, n=126). BI-RADS classifications differed significantly between MMG and DBT, especially in dense breasts. Of the BI-RADS 2 cases on MMG, 21.6% (n=64) were reclassified as BI-RADS 3 or 4 with DBT, while 6.3% (n=7) of BI-RADS 3 cases were downgraded. BI-RADS 5 categorization showed complete agreement between modalities, though 33% (n=5) of BI-RADS 4 cases were downgraded. At least one pathological finding was observed in 91.3% (n=475) of patients, most commonly nodular opacities (57.1%, n=297), vascular wall calcifications (13.3%, n=69), and coarse calcifications (5.4%, n=28). McNemar analysis revealed significant reclassification between BI-RADS 2 and 3 in Types C and D ($P<0.001$).

Conclusion: DBT significantly alters BI-RADS categorization in dense breasts. These findings support the potential role of DBT as a complementary tool in screening protocols, particularly for patients with heterogeneously dense or extremely dense breast tissue.

Keywords: Breast Tomosynthesis, Breast Imaging Reporting and Data System, Breast Density, Mammography, Breast Cancer Screening

Submitted: December 16, 2025 **Accepted:** February 1, 2026 **Published Online:** February 3, 2026

How to cite this article: Tekcan Şanlı DE, Aksu E, Şanlı AN, Turan B. The Effect of Digital Breast Tomosynthesis on BI-RADS Categorization in Different Breast Densities: A Retrospective Evaluation. *Eur Res J.* 2026;12(7):719-729. doi: 10.18621/eurj.1843405

Corresponding author: Bilal Turan, MD., Assist. Prof., Phone: +90 246 211 00 00, E-mail: bturan117@gmail.com

Available Online at https://dergipark.org.tr/en/pub/eurj

Breast cancer is the most common malignancy in women and a significant cause of morbidity and mortality worldwide [1]. Treatment success and survival rates significantly increase in breast cancers detected at an early stage, highlighting the importance of effective screening strategies [1, 2]. Mammography (MMG) is currently considered the most widely used imaging method for breast cancer screening and forms the basis of national screening programs in many countries [2]. However, the diagnostic efficacy of conventional mammography is significantly affected by breast tissue density [3]. Especially in women with dense breast tissue (Type C+D), fibroglandular structures can mask potential lesions, leading to false-negative results and delaying diagnosis [3]. This demonstrates that breast density is not merely an imaging characteristic but also an important factor influencing clinical decisions.

Digital breast tomosynthesis (DBT) is a technology developed to overcome these limitations, reducing tissue superposition by imaging breast tissue in layers [4]. DBT has been shown to make lesions, particularly those with ill-defined borders or those masked within the breast parenchyma, easier to detect [4]. Furthermore, synthetic mammograms created from tomosynthesis images have made it possible to obtain diagnostic information without additional radiation exposure [4].

Recent studies have shown that tomosynthesis provides higher diagnostic accuracy compared to conventional mammography, particularly in dense breast tissue [5]. However, the impact of this technology on the Breast Imaging Reporting and Data System (BI-RADS) classification, developed by the American College of Radiology (ACR) and widely used to guide clinical management, remains unclear. Whether the changes induced by tomosynthesis in BI-RADS categorization vary, particularly based on breast density, is an important research topic.

In this context, this study aimed to retrospectively evaluate the modifying effect of digital breast tomosynthesis on BI-RADS categories determined by conventional mammography in individuals with different breast densities. The contribution of tomosynthesis to the diagnostic process and the possibility of reclassifying suspicious lesions, especially in patients with dense breast structure, were comparatively investigated.

METHODS

Study Design and Ethical Approval

This study was designed as a retrospective observational study. Archived images of female patients who underwent MMG and simultaneous DBT for routine breast cancer screening or preliminary diagnosis of malignancy at our hospital between January 2023 and June 2025 were evaluated. Local ethics committee approval was obtained for the study (Gaziantep University, No: 2025/157), and all patient data were anonymized and handled in accordance with the principles of the Declaration of Helsinki.

Patient Selection

The study included female patients aged 35–85 years who underwent MMG for screening or diagnostic purposes and had MMG and DBT images available in the institutional archive. Inclusion criteria were technically adequate image quality and availability of a complete dataset for both modalities. During the study period, records meeting these criteria were retrospectively reviewed, and 1,200 patients were initially identified. Patients with prior breast surgery, previous chest radiotherapy, pregnancy or lactation, or incomplete/artifact-containing images were excluded (n=680). After applying these exclusion criteria, a final cohort of 520 patients was included in the analysis.

Imaging Protocol

All imaging studies were performed using the same device (Fujifilm Amulet Innovality, DBT system) and by the same technician according to a standard protocol. Conventional mammography images were acquired in the craniocaudal (CC) view, while digital breast tomosynthesis (DBT) was performed in the mediolateral oblique (MLO) view. The MLO view for mammographic evaluation was provided by synthetic MLO (sMLO) images automatically generated from the DBT data; therefore, no additional conventional full-field digital mammography MLO acquisition was performed. This allowed the mammographic assessment to be completed using a standard two-view approach (CC + MLO) without additional radiation exposure. This imaging strategy reflects current clinical practice, as DBT is most commonly performed in the MLO

position to provide the widest breast coverage with minimal radiation exposure [4, 6, 7]. Previous studies have demonstrated that the combination of DBT with synthesized two-dimensional images achieves diagnostic accuracy comparable to, or even superior to, conventional two-view full-field digital mammography (FFDM), while avoiding additional radiation exposure [8, 9]. Cost-effectiveness analyses have further supported the clinical adoption of DBT combined with synthetic mammography as a viable alternative to conventional FFDM [10]. Tomosynthesis scans were performed at 2-mm intervals with 15° narrow-angle projections, and technical parameters were automatically adjusted according to breast thickness and compression strength.

Mammographic and Tomosynthesis Evaluation

MMG images, consisting of conventional CC and sMLO views, were evaluated in consensus by two experienced radiologists (DETS, EA) using the ACR BI-RADS Mammography Lexicon as the reference standard [11]. The evaluations for MMG and DBT were not performed simultaneously. All MMG images were first reviewed in consensus by two experienced radiologists. Subsequently, in separate sessions held on different days, the same radiologists evaluated the DBT images. During DBT interpretation, the readers were blinded to their prior MMG assessments and to any available histopathological results. This approach was chosen to minimize recall bias and to ensure methodological rigor.

Patients were categorized as Type A-B-C-D based on mammographic breast density. According to the ACR criteria [11]. Based on the available imaging findings, the BI-RADS category was determined and recorded for each patient, first for MMG and then for DBT. Pathological findings (mass opacity, nodular opacity, asymmetric density, coarse calcification-vascular calcification, microcalcification, distortion) detected in cases other than BI-RADS 1, as well as the 2D maximum transverse diameters of these pathologies, were also noted. DBT findings were taken into account in lesion characterization and classification. In cases where more than one radiological finding was observed simultaneously, the characteristic that was decisive for malignancy was selected as the dominant finding.

In the study, benign findings most frequently observed included coarse calcifications, vascular wall calcifications (BI-RADS 2), and oval, circumscribed nodular opacities <1 cm in size with benign morphology (BI-RADS 2–3). In such cases, the most dominant imaging characteristic was recorded. For example, when both BI-RADS 2 and BI-RADS 3 features coexisted, the case was categorized as BI-RADS 3. When more than one BI-RADS 2 feature was present (e.g., coarse calcifications together with vascular wall calcifications), the case was classified as BI-RADS 2. Other benign descriptors such as intramammary lymph nodes with typical appearance, fat-containing lesions (such as lipomas, hamartomas), and secretory-type calcifications were also accepted as BI-RADS 2. Lesions with probably benign features such as focal asymmetries or circumscribed masses without suspicious morphology were assigned to BI-RADS 3, in line with the ACR BI-RADS lexicon [11].

In patients divided into four groups based on mammographic breast density, differences observed between the BI-RADS categories determined by MMG and DBT were analyzed both at the population level and within breast density subgroups.

Statistical Analysis

All statistical analyses were performed using SPSS (v25.0). For descriptive statistics, mean, standard deviation, minimum, and maximum values were calculated for numerical variables, and frequencies and percentage distributions were calculated for categorical variables. Patients were divided into four main groups based on breast density (Types A, B, C, and D). The agreement between BI-RADS categories determined by MMG and DBT was assessed separately for all cases and density subgroups. The McNemar test was used to evaluate differences in paired categorical data and to determine whether tomosynthesis caused reclassification. Statistical significance was set at $P<0.05$ in all analyses.

RESULTS

Demographic Data and Imaging Distributions

A total of 520 female patients with available digital MMG and DBT images were included in this

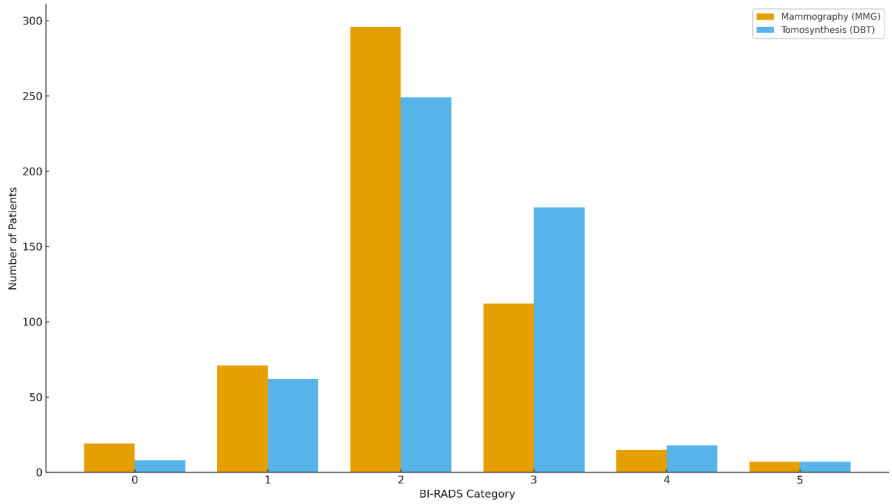

FIGURE 1. **BI-RADS category distribution chart comparing Mammography (MMG) and Tomosynthesis (DBT). BI-RADS, Breast Imaging Reporting and Data System.**

study. The mean age of the patients was 53.0±9.9 years. Based on mammographic breast density, 3.8% (n=20) of the patients were classified as Type A, 28.5% (n=148) as Type B, 43.5% (n=226) as Type C, and 24.2% (n=126) as Type D. The mean ages for the density groups were 52.3±4.9 for Type A, 57.2±10.2 for Type B, 54.6±9.9 for Type C, and 47.2 ± 6.3 for Type D, respectively. The Type D group was observed to be concentrated in the younger age group. Among all mammography images in the study, 13.0% (n=71) were reported as normal (BI-RADS 1), while at least one pathological radiological finding was detected in the remaining 91.3% (n=475). The most common pathological lesion type was nodular opacity, observed in 57.1% (n=297) of all cases. This was followed by vascular wall calcification (13.3%, n=69), coarse

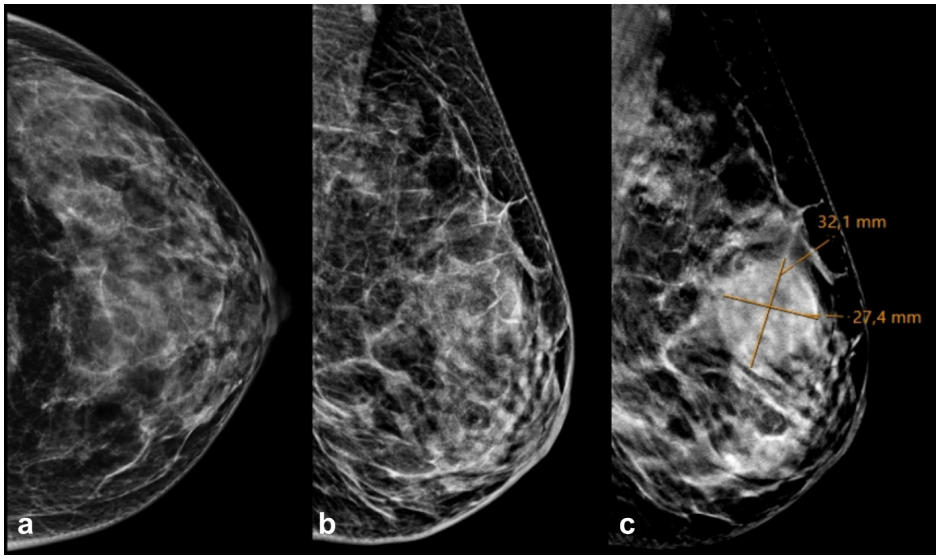

FIGURE 2. **In a 46-year-old female patient who underwent screening MMG, a regional asymmetric density increase was observed in the conventional MMG image at the 12 o'clock position in the left breast, while a 32×27 mm well-circumscribed mass opacity was detected at this level in the DBT images (BI-RADS 3) (evaluated in favor of fibroadenoma by ultrasound). MMG, Mammography; DBT, Tomosynthesis; BI-RADS, Breast Imaging Reporting and Data System.**

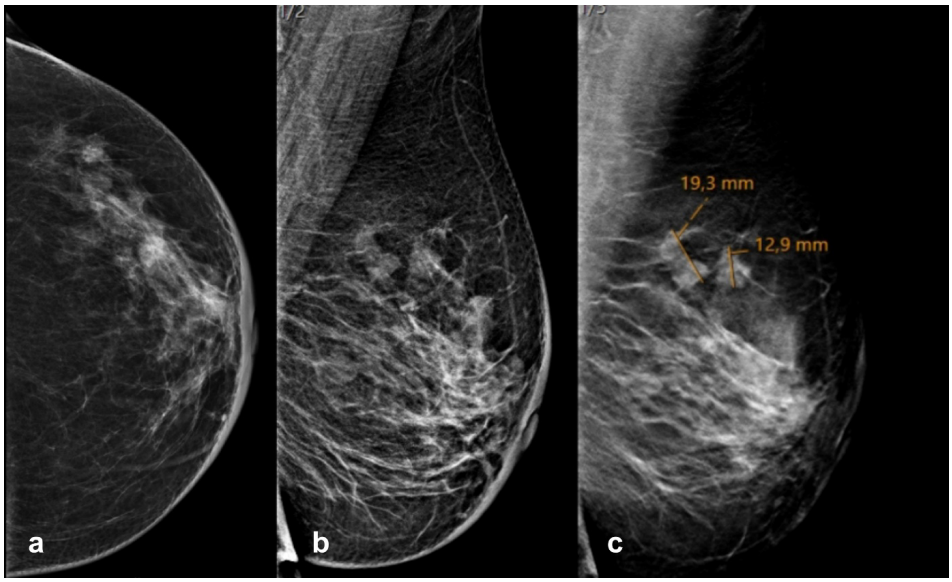

FIGURE 3. A 39-year-old woman presented with a palpable hardness in her left breast. A diagnostic MMG scan performed for a preliminary diagnosis of malignancy revealed adjacent millimetric nodular opacities with unclear contours in the upper outer quadrant of her left breast. The DBT image revealed two lesions with irregular contours (BI-RADS 4C). (Ultrasound assessed the lesions as highly suspicious for malignancy, and the biopsy result was reported as invasive ductal carcinoma, grade 3, Estrogen Receptor: 90%, Progesterone Receptor: 60%, and Ki-67 index: 20%). MMG, Mammography; DBT, Tomosynthesis; BI-RADS, Breast Imaging Reporting and Data System.

calcification (5.4%, n=28), mass density (4.8%, n=25), asymmetric density (4.6%, n=24), distortion (1.9%, n=10), and microcalcification (1.2%, n=6).

In the evaluation performed with conventional MMG, 3.7% (n=19) of all images were reported as BI-RADS 0; 13.7% (n=71) as BI-RADS 1; 56.9% (n=296) as BI-RADS 2; 21.5% (n=112) as BI-RADS 3; 2.9% (n=15) as BI-RADS 4 and 1.3% (n=7) as BI-RADS 5. In the re-evaluation performed with DBT, 1.5% (n=8) were classified as BI-RADS 0; 11.9% (n=62) as BI-RADS 1; 47.9% (n=249) as BI-RADS 2; 33.8% (n=176) as BI-RADS 3; 3.5% (n=18) as BI-RADS 4; and 1.3% (n=7) as BI-RADS 5 (Figures 1 to 4).

The number of patients with BI-RADS 4 (n=15) and BI-RADS 5 (n=7) by MMG was 22; with DBT (18/7) was 25. Pathology results of 22 of these patients were available in the system and a total of 17 patients were diagnosed as malignant histopathologically (BI-RADS 4: 10 malignant; BI-RADS 5: 7 malignant).

BI-RADS Distribution at the General Population Level

When the total cases included in the study were evaluated, differences were observed between the BI-RADS categories assigned on MMG and those after DBT. Reclassifications were particularly notable in BI-RADS 2 and 3. Of the 296 cases assessed as BI-RADS 2 on MMG, 61 (20.6%) were reclassified to BI-RADS 3 on DBT; additionally, 3 cases were upgraded to BI-RADS 4 and 8 were downgraded to BI-RADS 1. Of the 112 cases classified as BI-RADS 3 on MMG, 5 were upgraded to BI-RADS 4 and 5 were downgraded to BI-RADS 2, while 2 were reclassified as BI-RADS 1. Some cases initially categorized as BI-RADS 4 on MMG were reclassified to lower categories after DBT. These shifts were primarily attributable to improved depiction of lesion morphology on DBT, including clearer margin assessment and better visualization of subtle architectural distortion.

BI-RADS Comparison Based on Breast Density

Type A (Fatty Breast)

In this group, the effect of DBT on BI-RADS classification was quite limited. Of the 6 patients classified as BI-RADS 2 with MMG, only 1 was reclassified to BI-RADS 3 with DBT, and 3 patients were classified as BI-RADS 3 with both MMG and

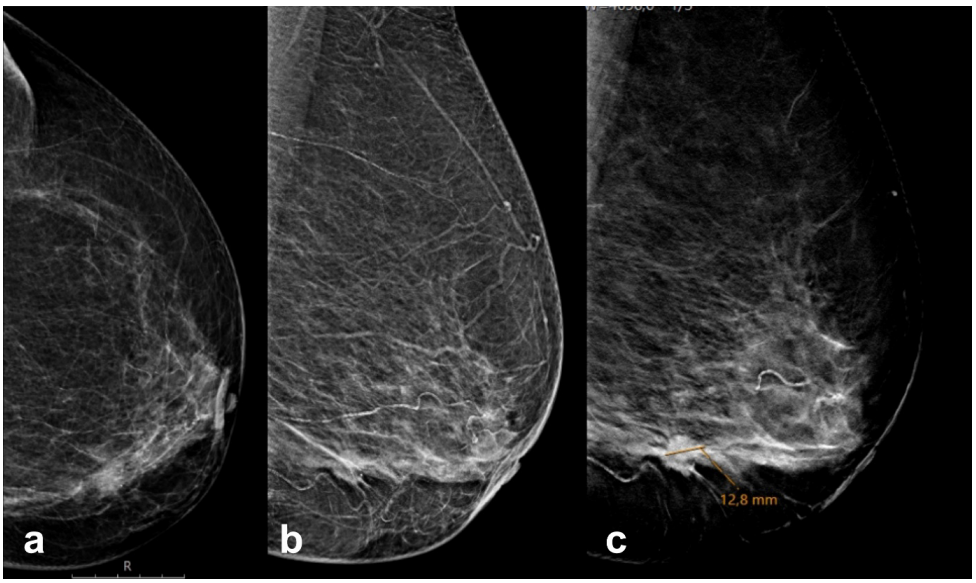

FIGURE 4. A 75-year-old woman presented with left breast hardness. A diagnostic MMG revealed linear asymmetric density accompanied by parenchymal distortion and retraction in the lower inner quadrant of the left breast, along with a millimetric nodular opacity with indistinct margins. On DBT, a 12.8 mm nodular opacity with irregular spiculated contours and highly suspicious for malignancy was clearly visible along with its contours (BI-RADS 5). (Ultrasound was also evaluated as BI-RADS 5, and the biopsy pathology report revealed invasive ductal carcinoma, grade: 3, Estrogen Receptor: 90%, Progesterone Receptor: 90%, and Ki-67 index: 10%). MMG, Mammography; DBT, Tomosynthesis; BI-RADS, Breast Imaging Reporting and Data System.

DBT. According to the McNemar test, this difference was not statistically significant (P=1.00). This group comprised a total of 20 patients.

Type B (Scattered fibroglandular densities)

In this group, 70 patients classified as BI-RADS 2 with MMG were reclassified with DBT, but 11 patients were upgraded to BI-RADS 3. Similarly, 2 patients classified as BI-RADS 3 were downgraded to BI-RADS 2 with DBT. The McNemar test found this difference to be statistically significant (P=0.022). The group included 148 patients.

TABLE 1. Cross-Tabulation of BI-RADS Categories Between Mammography and Digital Breast Tomosynthesis

MMG	BI-RADS 0	BI-RADS 1	BI-RADS 2	BI-RADS 3	BI-RADS 4	BI-RADS 5	Total (MMG)
DBT							
BI-RADS 0 (n=19)	0	8	1	10	0	0	19
BI-RADS 1 (n=71)	0	52	17	2	0	0	71
BI-RADS 2 (n=296)	0	8	224	61	3	0	296
BI-RADS 3 (n=112)	0	2	5	100	5	0	112
BI-RADS 4 (n=15)	0	0	2	3	10	0	15
BI-RADS 5 (n=7)	0	0	0	0	0	7	7

Values represent the number of patients reclassified between BI-RADS categories when comparing conventional mammography (MMG) and digital breast tomosynthesis (DBT). BI-RADS, Breast Imaging Reporting and Data System. Rows indicate the initial BI-RADS category based on MMG, and columns represent the final BI-RADS category after DBT.

Type C (Heterogeneous Dense Breast)

One of the most significant effects of DBT was observed in this group. Of the 103 patients with BI-RADS 2, 25 were upgraded to BI-RADS 3 with DBT, and 3 BI-RADS 3 patients were downgraded to BI-RADS 2 with DBT. The McNemar test showed that this difference was highly significant (P<0.001). There were a total of 226 patients in this density group.

Type D (Extremely Dense Breast)

The most significant reclassification rates were observed in this group. Of the 45 patients with BI-RADS 2 with MMG, 24 were upgraded to BI-RADS 3 with DBT. The 21 patients with BI-RADS 3 remained unchanged; no regression was observed. According to the McNemar test results, this difference was highly statistically significant (P<0.001). Type D group consists of 126 patients.

Re-Evaluation of BI-RADS 0 Cases

A total of 19 (3.5%) cases classified as BI-RADS 0 by MMG were referred for further evaluation due to incomplete evaluation or inability to clarify a suspicious finding. Re-evaluation with DBT resulted in a clear classification of all these cases. 10 (52.6%) cases were re-classified as BI-RADS 3, 8 (42.1%) as BI-RADS 1, and 1 (5.3%) as BI-RADS 2 (Table 1).

Re-Classification of BI-RADS 1 Cases

Of the 71 (13.7%) cases classified as BI-RADS 1 by MMG, re-evaluation with DBT resulted in only 52 (73.2%) retaining the BI-RADS 1 classification. Of the remaining cases, 17 (23.1%) were reclassified as BI-RADS 2 with DBT, and 2 (2.8%) as BI-RADS 3 (Table 1).

Reclassification of BI-RADS 2 Cases

Of the 296 (56.9%) cases identified as BI-RADS 2, 224 (75.7%) remained in the same category with DBT; 61 (20.6%) were reclassified as BI-RADS 3, and 3 (1.0%) were reclassified as BI-RADS 4 (Figure 5). Additionally, 8 (2.7%) cases were reclassified as BI-RADS 1 because previously identified benign findings on DBT were no longer visible or more clearly distinguished. Within this group, the proportion of cases in which the BI-RADS category changed with DBT was approximately 24% (Table 1).

Reclassification of BI-RADS 3 Cases

Of the 112 (21.5%) cases assessed as BI-RADS 3 by MMG, only 100 (89.3%) maintained the same category with DBT. Of the remaining 12 (10.7%) cases, 5 (4.5%) were reclassified to BI-RADS 2 due to lower suspicion, and 5 (4.5%) were reclassified to BI-RADS 4 due to increased suspicion (Figure 5).

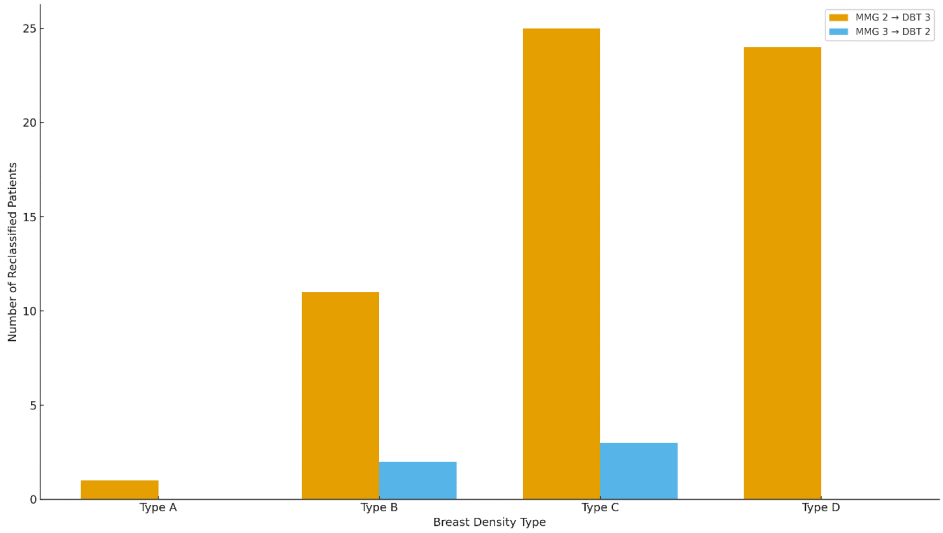

FIGURE 5. The bar chart illustrates the reclassification of patients between BI-RADS categories 2 and 3 based on breast density type when comparing conventional MMG with DBT. MMG, Mammography; DBT, Tomosynthesis; BI-RADS, Breast Imaging Reporting and Data System.

Additionally, 2 (1.8%) cases were assessed within normal limits by DBT and were reclassified as BI-RADS 1 (Table 1).

Reclassification of BI-RADS 4 Cases

A detailed examination of the 15 (2.9%) cases assessed as BI-RADS 4 by MMG, revealed that only 10 (66.7%) of these cases maintained the BI-RADS 4 category after reevaluation with DBT. Of the remaining 5 (33.3%) cases, 3 (20%) were reclassified to BI-RADS 3 by DBT, and 2 (13.3%) were reclassified to BI-RADS 2 (Table 1).

Consistency of BI-RADS 5 Cases

All 7 (1.3%) cases classified as BI-RADS 5 by MMG were also classified as BI-RADS 5 by DBT (Table 1).

DISCUSSION

This study aimed to evaluate the effect of DBT on BI-RADS categorization in different breast density types compared to conventional MMG. The results demonstrate that DBT significantly alters BI-RADS classification, particularly in women with dense breast tissue, and thus contributes significantly to the diagnostic process.

The overall findings of the study revealed a remarkable reclassification rate between BI-RADS 2 and 3 categories. More than 20% of patients assessed as BI-RADS 2 on MMG were upgraded to BI-RADS 3 or 4 with DBT. Similarly, some cases identified as BI-RADS 3 on MMG were downgraded to BI-RADS 2 as less suspicious, while others were upgraded to BI-RADS 4. This suggests that the ability of tomosynthesis to more fully delineate lesion morphology allows for more accurate classification of suspicious lesions. Approximately 27% of cases assessed as BI-RADS 1 were reclassified to higher categories such as BI-RADS 2 or 3, with new findings detected by tomosynthesis. This reclassification highlights the clinical importance of tomosynthesis, particularly in dense tissue that may cause a masking effect on mammography. Our findings demonstrate that DBT substantially modifies BI-RADS categorization, particularly through reclassification between BI-RADS 2 and 3, which were the most frequent and statistically significant shifts. Beyond this, DBT contributed to reducing the proportion of indeterminate examinations: all BI-RADS 0 cases were reassigned into definitive categories, and a proportion of BI-RADS 1 cases were reclassified into BI-RADS 2 or 3, underscoring its potential to decrease recalls and improve diagnostic clarity. Moreover, the bidirectional reclassification observed in BI-RADS 3 suggests that DBT reduces diagnostic uncertainty in this challenging intermediate category. Finally, the observation that some BI-RADS 4 cases were downgraded following DBT highlights its potential role in reducing unnecessary biopsies, a trend also supported by previous studies. Collectively, these results emphasize the clinical utility of DBT as a complementary tool to conventional MMG. The reclassification of BI-RADS categories in our study was primarily based on specific morphological features that could be better delineated with DBT compared to MMG. For example, several lesions initially classified as BI-RADS 3 on MMG were upgraded to BI-RADS 4 on DBT because margin characteristics such as microlobulation or irregular contours became more apparent in the tomosynthesis slices. Conversely, some BI-RADS 3 lesions were downgraded to BI-RADS 2 when DBT clarified their smooth, circumscribed margins and confirmed benign features. In addition, distortions that were equivocal on MMG could be identified more clearly with DBT, leading to upgrades. The histopathology results were available in 22 patients classified as BI-RADS 4 or 5, of which 17 were malignant. However, systematic follow-up was not part of the study design, which we acknowledge as a limitation. The upgrade and downgrade decisions in our study were mainly driven by margin assessment and the detection of subtle architectural distortions on DBT. These mechanisms have also been highlighted in previous studies. It was reported that DBT was particularly effective in reducing false positives by refining the reclassification between BI-RADS 3 and 4 categories [8]. Similarly, Rafferty *et al*. [12] demonstrated that DBT improves margin visibility and distortion detection, thereby supporting more reliable upgrade decisions. Our findings are consistent with these reports, underscoring that the morphological features better delineated by DBT play a pivotal role in BI-RADS category shifts.

In our cohort, microcalcifications were detected in only 1.1% of patients, which is lower than reported in some previous studies [13]. This relatively low proportion can be explained by the characteristics of our study population and the retrospective design. Most of the patients presented with nodular opacities or benign calcifications, such as vascular wall or coarse calcifications, while suspicious microcalcifications were less frequent. In addition, the technical specifications of the DBT system may have influenced our results. We used a narrow-angle DBT system, which is reported to provide relatively better visualization of microcalcifications but may be less advantageous than wide-angle systems for depicting architectural distortions and margin characteristics of masses [14]. Conversely, wide-angle DBT systems (≈40°) reduce structural noise and improve lesion conspicuity, particularly for spiculated or microlobulated masses and distortions. Therefore, while our narrow-angle system could theoretically favor microcalcification assessment, the overall low rate of microcalcifications in our study is more likely explained by patient population characteristics.

Analyses performed according to density subgroups also support this conclusion. While no significant difference was found between the BI-RADS categories determined by MMG and DBT in Type A and Type B (fatty and less dense breast) groups, the BI-RADS 2 ↔ 3 transitions were shown to be statistically significant in Type C and Type D (heterogeneously and extremely dense breast) groups ($P<0.001$). This finding highlights the issue of "decreased sensitivity of mammography in dense breast tissue," frequently emphasized in the literature, and suggests that tomosynthesis may compensate for this disadvantage [15]. Large population-based screening trials have similarly reported that DBT increases cancer detection rates and reduces recall rates in women with dense breasts [16, 17].

On the other hand, analyses conducted for the highly suspicious categories BI-RADS 4 and 5 revealed a different pattern. All cases assessed as BI-RADS 5 by MMG were also classified in the same way by tomosynthesis; some cases with BI-RADS 4 were downgraded to lower categories by DBT. This demonstrates that tomosynthesis can serve as a confirmatory tool, particularly in highly suspicious cases, and can also contribute to the prevention of unnecessary further examinations.

This study also demonstrates that tomosynthesis offers multifaceted benefits in terms of both diagnostic accuracy and its impact on clinical decision-making. The more precise evaluation of some indeterminate lesions, particularly those in the BI-RADS 3 category, with DBT provides valuable contributions to clinical decision-making by preventing unnecessary short-term follow-up and potentially missed malignant lesions.

These findings are consistent with the existing literature. Previous studies have also reported that DBT offers a higher lesion detection rate and fewer false-negative results compared to conventional MMG, particularly in dense breasts [6, 18, 19]. Additionally, there are results indicating that tomosynthesis can reduce false-positive rates, thus reducing unnecessary biopsies and anxiety [10, 20-22].

Strengths and Limitations

This study has several limitations. First, its retrospective and single-center design may limit the generalizability of the findings. Mammographic evaluation was performed using conventional CC images combined with synthetic MLO images, allowing a two-view assessment comparable to routine MMG; however, DBT acquisition was limited to the MLO projection. As CC DBT images were not available, lesions located outside the optimal MLO coverage may have been missed, potentially introducing bias. Nevertheless, this approach reflects routine clinical practice aimed at minimizing radiation exposure, as MLO DBT provides the widest breast coverage with a single projection.

Second, interobserver variability could not be assessed because all evaluations were performed in consensus by two experienced radiologists, precluding statistical agreement analysis. In addition, although the overall cohort size was relatively large, the limited number of patients in BI-RADS 4 and 5 categories restricts the reliability of subgroup analyses for high-suspicion lesions.

Another limitation is the lack of systematic histopathological confirmation for most cases. However, as the majority of patients had benign imaging features, biopsies were not clinically

indicated. Importantly, the primary aim of this study was not to assess diagnostic accuracy or malignancy prediction, but to evaluate the effect of DBT on BI-RADS categorization across different breast density groups.

Finally, DBT was available for all patients included in the retrospective dataset, including those with fatty breasts. While this allowed for uniform comparison between MMG and DBT across all breast density categories, it may limit the generalizability of our findings to real-world screening settings where DBT is often selectively applied, particularly in women with dense breasts. Despite these limitations, our findings are consistent with large population-based screening trials, such as the STORM and Oslo Tomosynthesis Screening Trials, which have demonstrated the added value of DBT, especially in dense breast tissue [16, 17].

CONCLUSION

This study demonstrated that DBT significantly influences BI-RADS categorization, particularly in women with dense breast tissue. DBT enabled both up- and down-classification of lesions, especially between BI-RADS 2 and 3, contributing to improved lesion characterization. The consistent performance in high-risk categories (BI-RADS 5) and enhanced detection of subtle findings further support its diagnostic value. These results support the potential role of DBT as a complementary tool to conventional MMG in routine breast cancer screening.

Ethics Approval and Consent to Participate

This study was approved by the Gaziantep University Non-Interventional Clinical Research Ethics Committee (Decision No: 2025/157; date: 02.07.2025). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Informed consent was not required in this study because this is a retrospective study.

Data Availability

All data generated or analyzed during this study are included in this published article. The data that support the findings of this study are available on request from the corresponding author, upon reasonable request.

Authors' Contribution

Study Conception: DETŞ, BT; Study Design: BT, DETŞ; Supervision: ANŞ, BT; Funding: N/A; Materials: DETŞ, EA; Data Collection and/or Processing: EA; Statistical Analysis and/or Data Interpretation: ANŞ, BT; Literature Review: ANŞ, DETŞ; Manuscript Preparation: BT, DETŞ; and Critical Review: EA, DETŞ.

Conflict of Interest

The author(s) disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The author(s) disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgments

The authors have no acknowledgments to declare.

Generative Artificial Intelligence Statement

The authors declare that no artificial intelligence-based tools or applications were used during the preparation of this manuscript. All content of the study was produced by the authors in accordance with scientific research methods and academic ethical principles.

Editor's Note

All statements made in this article are solely those of the authors and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Arnold M, Morgan E, Rumgay H, et al. Current and future burden of breast cancer: Global statistics for 2020 and 2040. Breast. 2022;66:15-23. doi: 10.1016/j.breast.2022.08.010.
2. Ren W, Chen M, Qiao Y, Zhao F. Global guidelines for breast cancer screening: A systematic review. Breast. 2022;64:85-99. doi: 10.1016/j.breast.2022.04.003.

3. Brown AL, Vijapura C, Patel M, De La Cruz A, Wahab R. Breast Cancer in Dense Breasts: Detection Challenges and Supplemental Screening Opportunities. Radiographics. 2023;43(10):e230024. doi: 10.1148/rg.230024.
4. Gao Y, Moy L, Heller SL. Digital Breast Tomosynthesis: Update on Technology, Evidence, and Clinical Practice. Radiographics. 2021;41(2):321-337. doi: 10.1148/rg.2021200101.
5. Raichand S, Blaya-Novakova V, Berber S, Livingstone A, Noguchi N, Houssami N. Digital breast tomosynthesis for breast cancer diagnosis in women with dense breasts and additional breast cancer risk factors: A systematic review. Breast. 2024;77:103767. doi: 10.1016/j.breast.2024.103767.
6. Chae EY, Kim HH, Cha JH, Shin HJ, Choi WJ. Detection and characterization of breast lesions in a selective diagnostic population: diagnostic accuracy study for comparison between one-view digital breast tomosynthesis and two-view full-field digital mammography. Br J Radiol. 2016;89(1062):20150743. doi: 10.1259/bjr.20150743.
7. Houssami N, Skaane P. Overview of the evidence on digital breast tomosynthesis in breast cancer detection. Breast. 2013;22(2):101-108. doi: 10.1016/j.breast.2013.01.017.
8. Gilbert FJ, Tucker L, Young KC. Digital breast tomosynthesis (DBT): a review of the evidence for use as a screening tool. Clin Radiol. 2016;71(2):141-150. doi: 10.1016/j.crad.2015.11.008.
9. Caumo F, Zorzi M, Brunelli S, et al. Digital Breast Tomosynthesis with Synthesized Two-Dimensional Images versus Full-Field Digital Mammography for Population Screening: Outcomes from the Verona Screening Program. Radiology. 2018;287(1):37-46. doi: 10.1148/radiol.2017170745.
10. Couto HL, Gargano LP, de Oliveira VM, et al. Cost-Effectiveness Analysis of Digital Breast Tomosynthesis Added to Synthetic Mammography in Breast Cancer Screening in Brazil. Pharmacoecon Open. 2024;8(3):403-416. doi: 10.1007/s41669-023-00470-7.
11. American College of Radiology (n.d.) BI-RADS Reporting System. Available from: https://www.acr.org/Clinical-Resources/Clinical-Tools-and-Reference/Reporting-and-Data-Systems/BI-RADS. Accessed 27 Jul 2025
12. Rafferty EA, Park JM, Philpotts LE, et al. Diagnostic accuracy and recall rates for digital mammography and digital mammography combined with one-view and two-view tomosynthesis: results of an enriched reader study. AJR Am J Roentgenol. 2014;202(2):273-281. doi: 10.2214/AJR.13.11240.
13. Rizuana IH, Leong MH, Tan GC, Isa ZM. Association Between Microcalcification Patterns in Mammography and Breast Tumors in Comparison to Histopathological Examinations. Diagnostics (Basel). 2025;15(13):1687. doi: 10.3390/diagnostics15131687.
14. Huang H, Scaduto D, Plaunova A, Rinaldi K, Fisher PR, Zhao W. Comparison of lesion detection and conspicuity between narrow-angle and wide-angle digital breast tomosynthesis for dense and non-dense breasts. J Med Imaging (Bellingham). 2023;10(Suppl 2):S22407. doi: 10.1117/1.JMI.10.S2.S22407.
15. Lynge E, Vejborg I, Andersen Z, von Euler-Chelpin M, Napolitano G. Mammographic Density and Screening Sensitivity, Breast Cancer Incidence and Associated Risk Factors in Danish Breast Cancer Screening. J Clin Med. 2019;8(11):2021. doi: 10.3390/jcm8112021.
16. Ciatto S, Houssami N, Bernardi D, et al. Integration of 3D digital mammography with tomosynthesis for population breast-cancer screening (STORM): a prospective comparison study. Lancet Oncol. 2013;14(7):583-589. doi: 10.1016/S1470-2045(13)70134-7.
17. Skaane P, Bandos AI, Gullien R, et al. Comparison of digital mammography alone and digital mammography plus tomosynthesis in a population-based screening program. Radiology. 2013;267(1):47-56. doi: 10.1148/radiol.12121373.
18. Kassis I, Lederman D, Ben-Arie G, Giladi Rosenthal M, Shelef I, Zigel Y. Detection of breast cancer in digital breast tomosynthesis with vision transformers. Sci Rep. 2024;14(1):22149. doi: 10.1038/s41598-024-72707-2.
19. Li J, Zhang H, Jiang H, et al. Diagnostic Performance of Digital Breast Tomosynthesis for Breast Suspicious Calcifications From Various Populations: A Comparison With Full-field Digital Mammography. Comput Struct Biotechnol J. 2018;17:82-89. doi: 10.1016/j.csbj.2018.12.004.
20. Mathew B. Effectiveness of psychological intervention package on anxiety and wellness level among patients with anxiety disorders. J Family Med Prim Care. 2022;11(11):6704-6713. doi: 10.4103/jfmpc.jfmpc_561_21.
21. Hadadi I, Clarke J, Rae W, McEntee M, Vincent W, Ekpo E. Reducing Unnecessary Biopsies Using Digital Breast Tomosynthesis and Ultrasound in Dense and Nondense Breasts. Curr Oncol. 2022;29(8):5508-5516. doi: 10.3390/curroncol29080435.
22. Ho TH, Bissell MCS, Kerlikowske K, et al. Cumulative Probability of False-Positive Results After 10 Years of Screening With Digital Breast Tomosynthesis vs Digital Mammography. JAMA Netw Open. 2022;5(3):e222440. doi: 10.1001/jamanetworkopen.2022.2440.

The European Research Journal 2026;12(7):730-738 Research Article

DOI: https://doi.org/10.18621/eurj.1839146 *Cardiovascular Surgery*

Comparative Analysis of Large Language Models in Hemodialysis Vascular Access: ChatGPT-5, Gemini-2.5, and DeepSeek-V3

Muhammet Hüseyin Erkan[1], Ömer Faruk Rahman[2], Abdullah Güner[3], Fevzi Ayyıldız[4], Emin Barbarus[1]

[1]Department of Cardiovascular Surgery, Balıkesir University, Faculty of Medicine, Balıkesir, Türkiye; [2]Department of Cardiovascular Surgery, İzmir Bakırçay University, Faculty of Medicine, İzmir, Türkiye; [3]Department of Cardiovascular Surgery, Konya City Hospital, Konya, Türkiye; [4]Department of Cardiovascular Surgery, Afyonkarahisar State Hospital, Afyonkarahisar, Türkiye

Abstract:

Objective: Vascular access is a key component of effective hemodialysis treatment. Patient education regarding selection, complications, and daily care remains challenging. Recently, large language models (LLMs) have been explored as supportive tools in this field. This study compared three widely used LLMs (ChatGPT-5, Gemini-2.5, and DeepSeek-V3) in addressing patient-centered questions on hemodialysis vascular access.

Methods: Twenty-five frequently asked patient questions were compiled from literature, educational materials, and expert input. Each question was submitted to the three LLMs in standardized sessions, and answers were anonymized. Four cardiovascular surgeons independently evaluated responses in a blinded manner using 5-point Likert scales for accuracy, clarity, and scientificity. Statistical analyses compared model performances.

Results: DeepSeek-V3 provided significantly responses with a higher word count and achieved higher scientific depth scores compared to ChatGPT-5 and Gemini-2.5 ($P<0.01$). Accuracy scores showed ChatGPT-5 demonstrated significantly lower than both Gemini-2.5 and DeepSeek-V3 ($P<0.001$). No significant differences were observed among models regarding clarity. Overall, DeepSeek achieved the highest mean scores across all criteria.

Conclusion: DeepSeek-V3 demonstrated superior scientific depth and overall reliability, whereas ChatGPT-5 showed relative weaknesses in accuracy. The comparable clarity across models highlights LLMs' potential as supportive tools for patient education. Future studies should further validate these tools in real-world clinical settings and define appropriate safeguards for their use.

Keywords: Large Language Models, Artificial Intelligence, Patient Education, Hemodialysis, Vascular Access

Chronic kidney disease (CKD) is a global public health problem with increasing prevalence and incidence, significantly reducing quality of life and closely associated with mortality [1]. Hemodialysis is one of the most commonly used renal replacement therapies for patients with end-stage renal disease. Adequate and reliable vascular access is vital for the effective implementation of hemodialysis. Arteriovenous fistula (AVF), arteriovenous graft (AVG), and central venous catheter (CVC) are the

Submitted: December 9, 2025 **Accepted:** January 15, 2026 **Published Online:** January 20, 2026

How to cite this article: Erkan MH, Rahman ÖF, Güner A, Ayyıldız F, Barbarus E. Comparative Analysis of Large Language Models in Hemodialysis Vascular Access: ChatGPT-5, Gemini-2.5, and DeepSeek-V3. *Eur Res J.* 2026;12(7):730-738. doi: 10.18621/eurj.1839146

Corresponding author: Muhammet Hüseyin Erkan, MD., Assist. Prof., Phone: +90 266 612 14 61, E-mail: mhuseyinerkan@gmail.com

Available Online at https://www.eurj.org.tr

main vascular access routes used today. International guidelines recommend AVF creation as the first option because it provides the highest possible patency, the lowest complication rate, and the best patient survival [2, 3].

However, patients may experience uncertainty and anxiety across a wide range of issues, from the choice of access type to the postoperative care process, from potential complications to the organization of daily life activities. Patients with insufficient knowledge may have difficulty recognizing clinical symptoms requiring medical intervention and, as a result, may delay seeking healthcare [4]. Furthermore, it has been reported that CKD patients with low health literacy have significantly lower quality of life parameters related to physical and mental health [5]. The literature shows that patients who can communicate effectively with healthcare professionals, evaluate reliable information sources, and understand and process health information have a higher prevalence of AVF use compared to catheters [6].

Traditionally, this information process has been carried out face-to-face by nephrologists, cardiovascular surgeons, and dialysis nurses. However, the intensity of healthcare systems, time constraints, and the large number of patients can make it difficult to allocate sufficient time to each patient. To fill this gap, patients are increasingly turning to the internet and digital resources. Revolutionary developments in artificial intelligence (AI) in recent years suggest that tools such as Large Language Models (LLMs) could play a potential role in patient education. Thanks to their natural language processing capabilities, these models have the potential to explain complex medical topics in language appropriate for patients. A study on the use of AI in nephrology reported that nephrologists who can integrate AI into their work can provide better care to their patients [7].

However, there are significant concerns regarding the quality, accuracy, and reliability of the content produced by these models. LLMs may reflect errors or biases in the datasets they were trained on, provide outdated information, or generate statements that sound plausible but are entirely unrealistic, known as “hallucinations” (8). When it comes to medical information, the risk that such errors could potentially harm patients necessitates a rigorous evaluation of these tools' performance. The current literature includes studies measuring the knowledge level of ChatGPT and other LLMs across various medical disciplines [9, 10]. However, no comprehensive study has been identified that provides a comparative evaluation of these models within a specific, multidisciplinary field such as hemodialysis vascular access, where patient education is crucial.

Therefore, access to accurate, understandable, and evidence-based information may play a pivotal role in patients’ decision-making regarding hemodialysis vascular access. Based on this premise, the aim of the present study is to evaluate the responses provided by three popular LLMs (ChatGPT-5, Gemini-2.5, and DeepSeek-V3) to frequently asked questions by patients about hemodialysis vascular access; evaluating them using a blinded method by an expert panel based on the criteria of accuracy, clarity, and scientific depth.

METHODS

The methodology of this study was designed to compare the performance of three LLMs in answering patient questions about hemodialysis vascular access. The study was based on a panel of expert physicians evaluating the responses obtained from the models using a blinded method and comparing these responses in terms of objective metrics. The models evaluated were ChatGPT-5 (OpenAI, Microsoft Corporation, San Francisco, CA, USA, 2025), Gemini-2.5 (Google, Mountain View, CA, USA, 2025), and DeepSeek-V3 (DeepSeek, Beijing, CHINA, 2025) versions, which were accessed via their public interfaces on August 21, 2025.

The question set consists of 25 questions covering a variety of topics, including preparation for fistula surgery, fistula maturation, daily life and protection, fistula complications and their management, and catheter care. The identification of frequently asked questions was carried out through a systematic evaluation of widely used online platforms, including Google Trends, YouTube search recommendations, patient support forums, and official web portals of relevant healthcare organizations. Furthermore, questions recurrently reported by patients in daily clinical practice were included based on the investigators’ observations.

During the data collection process, each question was presented with a standardized prompt for each model. As an important methodological precaution, the browser cache and cookies were cleared for each new question, and a new chat session was always opened to prevent previous interactions from influencing the model output. All responses received were recorded in plain text, stripped of formatting, along with the relevant question and model ID.

To evaluate the collected responses, 25 responses from each model were randomly labeled with letters (Model X, Y, Z) and compiled into separate booklets. No information revealing the model ID was included anywhere in the booklets. The evaluation was conducted by a panel of four cardiovascular surgery specialists. A calibration meeting was held prior to the evaluation to ensure that the panelists interpreted the scales consistently. A minimum of 72 hours was allowed between the evaluation of one booklet and the distribution of the next to clear the evaluators' memory (washout). Each response was scored by the panelists on a 5-point Likert scale according to three main criteria: accuracy of medical content, clarity and comprehensibility of information for patients, and scientific depth and scope of information provided.

Statistical Analysis

All data were analyzed using IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp., Armonk, NY, USA) software. The normality of data distribution was assessed using the Shapiro–Wilk test. Differences in response scores between the three models were examined using repeated measures analysis of variance for normally distributed data and the Friedman test for non-normally distributed data. Results for normally distributed data were reported as mean ± standard deviation, while results for non-normally distributed data were reported as median (minimum–maximum) values. A statistical significance level of $P<0.05$ was accepted for all analyses.

RESULTS

In terms of response length, defined as the total word count of each generated answer, statistically significant differences were observed among ChatGPT-5, Gemini-2.5, and DeepSeek-V3. DeepSeek-V3 produced the longest responses, with a mean word count of 439.12, followed by Gemini-2.5 (257.08) and ChatGPT-5 (149.60). These differences were statistically significant ($P<0.001$). Pairwise comparisons demonstrated that DeepSeek-V3 responses contained a significantly higher number of words than those of both Gemini-2.5 and ChatGPT-5 ($P<0.01$ for both comparisons), and that Gemini-2.5 responses were also significantly longer than those of ChatGPT-5 ($P<0.001$). Detailed data on response length are presented in Table 1.

For Rater 1, a statistically significant difference was found between ChatGPT-5, Gemini-2.5, and DeepSeek-V3 in the scientificity category (Figure 1). In pairwise comparisons, this difference was observed between ChatGPT-5 and DeepSeek-V3 ($P=0.022$). DeepSeek's median value was found to be significantly higher than ChatGPT-5's (5 and 4, respectively; $P=0.04$). No statistically significant differences were found in the other categories evaluated for R1 ($P=0.294$).

No statistically significant differences were found among the three models in the accuracy category for Rater 2 ($P=0.312$). Statistically significant differences were found between ChatGPT-5, Gemini-2.5, and DeepSeek-V3 in the scientificity category for R2 ($P<0.001$). In the pairwise comparisons, DeepSeek scored significantly higher than ChatGPT-5 ($P<0.001$) and was also significantly superior to Gemini-2.5 ($P<0.001$). No significant difference was found among the three models in the clarity category ($P=0.113$).

For Rater 3, statistically significant differences were found among the three models in the accuracy category ($P<0.001$). In the pairwise comparisons, ChatGPT-5 scored significantly lower than both Gemini-2.5 and DeepSeek-V3 ($P<0.001$ for both comparisons). In contrast, no significant difference was found between Gemini-2.5 and DeepSeek-V3 ($P=0.832$). For R3, a statistically significant difference was found among the three models in the scientificity category ($P<0.001$). In pairwise comparisons, ChatGPT-5 scored significantly lower than both Gemini-2.5 ($P=0.007$) and DeepSeek-V3 ($P<0.001$). In contrast, no significant difference was found between Gemini-2.5 and DeepSeek-V3 ($P=0.437$). No statistically significant difference was found among the models in the clarity category for R3 ($P=1.000$).

TABLE 1. Comparison of Response Length (Measured as Word Count) Among ChatGPT-5, Gemini-2.5, and DeepSeek-V3

Questions	ChatGPT-5	Gemini-2.5	DeepSeek-V3
1. Why is it necessary to establish a vascular access for hemodialysis?	171	321	433
2. What is the difference between an arteriovenous fistula and a catheter? Which one is safer?	180	366	524
3. Which vascular access option is more suitable for me for hemodialysis?	243	438	570
4. What tests are performed before fistula surgery?	171	285	574
5. Is fistula surgery difficult? How long does it take?	116	143	482
6. Do I need to do any special preparation before fistula surgery?	195	214	410
7. How long after fistula surgery will it be ready for use?	102	185	333
8. What does it mean for a fistula to "mature"? How long does it take?	160	242	419
9. What should I do after surgery to protect my fistula?	194	283	489
10. How can I tell if my fistula is working?	143	250	449
11. Can I use my fistula arm? Can I lift heavy objects?	123	215	374
12. Can blood pressure measurements or blood draws be performed on my fistula arm?	73	181	242
13. What should I be careful about in daily life to avoid damaging my fistula?	177	237	408
14. What should I do if my fistula becomes swollen, painful, or bruised?	128	243	489
15. What happens if my fistula becomes blocked? Can it be reopened?	103	302	347
16. How can I tell if a clot has formed in my fistula?	120	201	445
17. Is it dangerous if a balloon (aneurysm) develops in my fistula?	155	250	463
18. How will I undergo dialysis if my fistula does not function?	136	247	435
19. Why is a catheter inserted, and why is it preferred over a fistula?	158	355	548
20. How can I tell if my dialysis catheter is infected?	125	317	287
21. Can I bathe with my dialysis catheter, and what should I pay attention to in daily care?	173	224	404
22. What should I do if my dialysis catheter becomes blocked? Does it need to be replaced?	118	292	393
23. What is a fistula graft, and in which cases is it preferred?	206	276	521
24. If my fistula fails, can I have another surgery?	114	190	441
25. If my vascular access is no longer viable, what is the last resort for dialysis?	156	170	498
Mean±SD	149.6±38.68	257.08±68.27	439.12±82.58
P-value		**<0.001**	
F		517.55	

SD, standard deviation; F, One-way Anova test statistic. Statistically significant P-value is shown in bold.

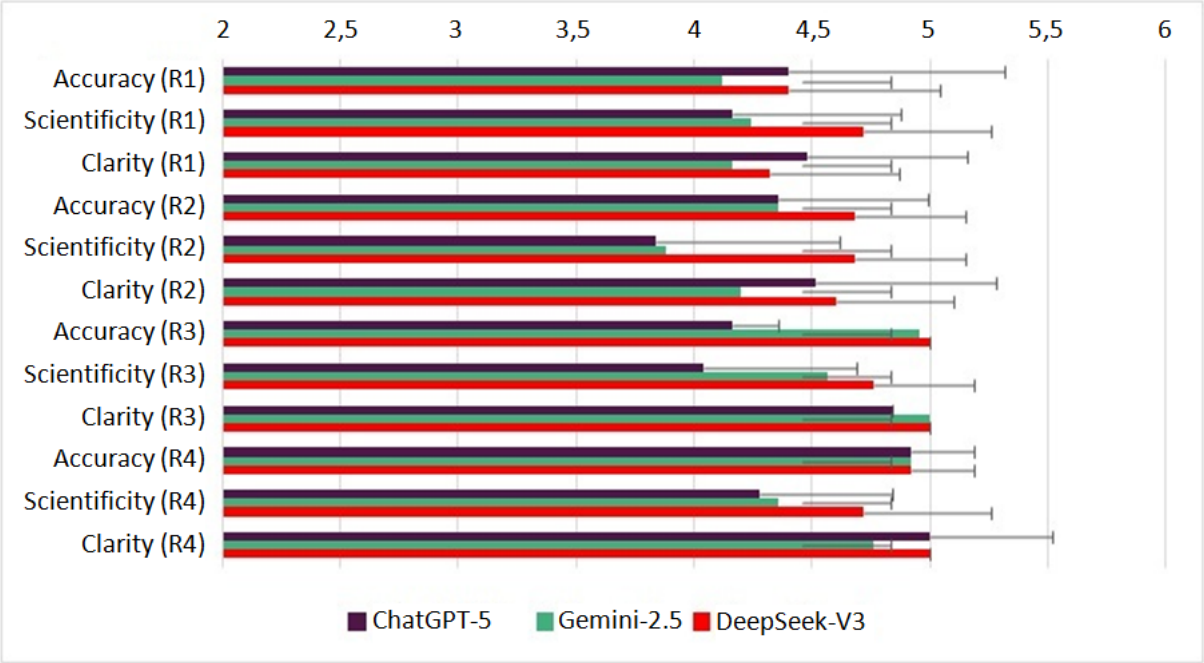

FIGURE 1. Comparison of ChatGPT, Gemini, and DeepSeek's scores in the accuracy, scientificity, and clarity categories using a bar chart.

TABLE 2. Comparison of ChatGPT-5, Gemini-2.5, and DeepSeek-V3's Accuracy, Scientificity, and Clarity Categories and Overall Average Scores

	ChatGPT-5	Gemini-2.5	DeepSeek-V3	stat	P-value
Accuracy (R1)	5 (3-5)	4 (2-5)	4 (3-5)	$\chi2$=2.452	0.294
Scientificity (R1)	4 (3-5)	4 (3-5)	5 (3-5)	$\chi2$=7.600	**0.022**
Clarity (R1)	5 (3-5)	4 (3-5)	4 (3-5)	$\chi2$=2.233	0.327
Accuracy (R2)	4 (4-5)	4 (3-5)	5 (4-5)	$\chi2$=6.500	0.312
Scientificity (R2)	4 (3-5)	4 (3-5)	5 (4-5)	$\chi2$=21.816	**<0.001**
Clarity (R2)	5 (3-5)	4 (3-5)	5 (4-5)	$\chi2$=4.353	0.113
Accuracy (R3)	4 (4-5)	5 (4-5)	5 (5-5)	$\chi2$=38.273	**<0.001**
Scientificity (R3)	4 (3-5)	5 (3-5)	5 (4-5)	$\chi2$=18.620	**<0.001**
Clarity (R3)	5 (4-5)	5 (5-5)	5 (5-5)	$\chi2$=8.000	1.0
Accuracy (R4)	5 (4-5)	5 (4-5)	5 (4-5)	$\chi2$=0	1.0
Scientificity (R4)	4 (3-5)	4 (3-5)	5 (3-5)	$\chi2$=9.864	**0.007**
Clarity (R4)	5 (5-5)	5 (3-5)	5 (5-5)	$\chi2$=10.000	0.051
Accuracy (average)	4.5 (4-5)	4.75 (3.5-5)	4.75 (4.25-5)	$\chi2$=18.074	**<0.001**
Scientificity (average)	4 (3.5-4.75)	4.25 (3.25-5)	4.75 (4-5)	$\chi2$=22.422	**<0.001**
Clarity (average)	4.75 (4.25-5)	4.5 (3.75-5)	4.75 (4.5-5)	$\chi2$=6.000	0.05
Three categories (average)	4.41±0.21	4.46±0.34	4.73±0.17	F=34.811	**<0.01**

$\chi2$, Friedman test statistic; F, one-way Anova test statistic; SD, standard deviation; Min, minimum; Max, maximum; R, rater (cardiovascular surgery specialist).

Statistically significant P-values are shown in bold.

No statistically significant difference was found among the models in the accuracy category for Rater 4 (P=1.000). A statistically significant difference was found between the models in the scientificity category for R4 (P=0.007). In the pairwise comparisons, this difference was observed between DeepSeek-V3 and ChatGPT-5 (P=0.02), and DeepSeek's scores were found to be significantly higher. No statistically significant difference was found between the models in the clarity category for R4 (P=0.051).

When the models were compared in terms of average accuracy scores, a statistically significant difference was found (P<0.001). In pairwise comparisons, ChatGPT-5 scored significantly lower than both Gemini-2.5 (P=0.034) and DeepSeek-V3 (P<0.001). In contrast, no significant difference was found between Gemini-2.5and DeepSeek-V3 (Table 2).

When comparing the models in terms of average scientificity scores, a statistically significant difference was found (P<0.01). In pairwise comparisons, DeepSeek-V3 scored significantly higher than both Gemini-2.5 (P=0.002) and ChatGPT-5 (P<0.001). No statistically significant difference was found between the models in terms of average clarity scores (P=0.051).

When the average scores for the three categories were compared (Figure 2), a statistically significant difference was found between ChatGPT-5, Gemini, and DeepSeek-V3 (P<0.001). DeepSeek-V3's average scores were found to be significantly higher than both Gemini-2.5 and ChatGPT-5 (4.73 compared to 4.46 and 4.41; P<0.01). In contrast, no significant difference was found between Gemini-2.5 and ChatGPT-5 in terms of average scores (P=1.000).

DISCUSSION

This is among the first studies to the performance of three LLMs (ChatGPT-5, Gemini-2.5, and DeepSeek-V3) in a specific medical field where patient education is critically important, such as hemodialysis vascular access, against expert evaluation. The findings reveal that, overall, the DeepSeek-V3 model demonstrated a statistically significant superiority over ChatGPT-5 and Gemini-2.5 in terms of accuracy, scientificity, and overall average scores. However, no significant difference was found among the models in the clarity category. This suggests that DeepSeek-V3's detailed responses do not reduce comprehensibility.

Previous studies evaluating LLMs in patient education across different medical domains have reported generally high performance in terms of clarity and patient-oriented explanations, with more variable

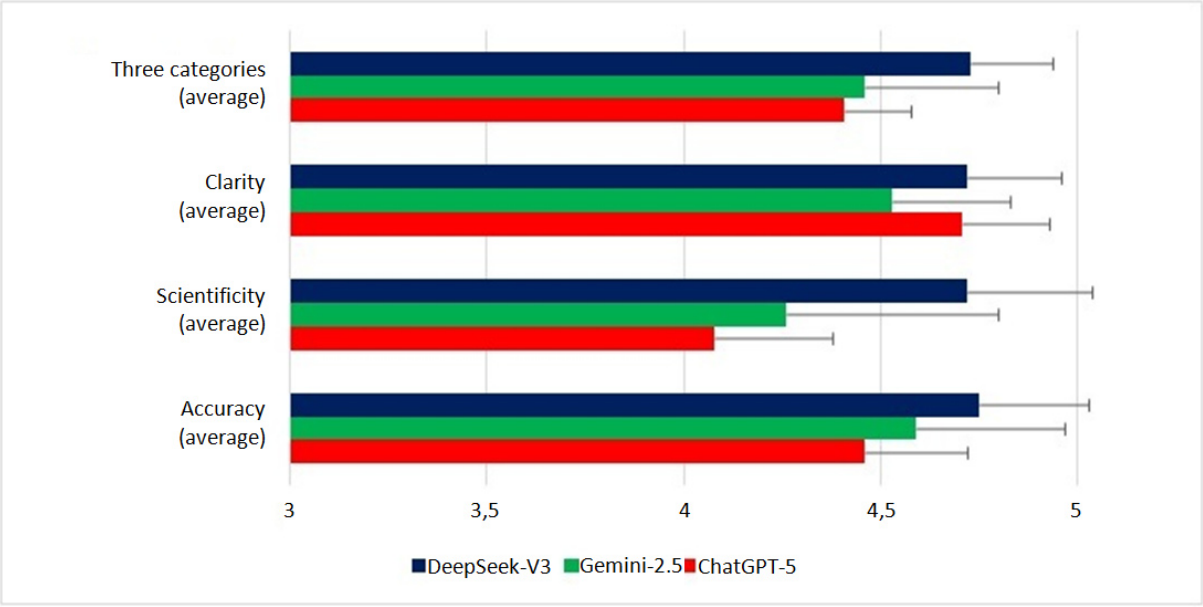

FIGURE 2. Comparison of ChatGPT, Gemini, and DeepSeek's accuracy, scientificity, clarity, and overall average scores using a bar chart.

results regarding scientific depth and accuracy. The literature has shown that LLMs can effectively communicate medical information, but performance can vary significantly depending on the clinical context and evaluation criteria [9, 10]. Consistent with these findings, our results indicate that although all evaluated models performed similarly in terms of clarity, notable differences were observed in scientific depth and accuracy, particularly in a multidisciplinary and procedure-specific field such as hemodialysis vascular access.

The most striking finding of our study is that DeepSeek-V3 scored significantly higher than the other two models on the scientificity criterion. This may be related to the scope, recency, and quality of the medical literature in the model's dataset. In particular, the evaluations by Rater 2 and Rater 3 showed that DeepSeek-V3 tends to present medical content in a more in-depth and evidence-based manner. This finding is extremely important when considering the role of LLMs in providing medical information, as it is essential to maintain scientific accuracy as well as comprehensibility for patients [11]. In this study, DeepSeek-V3's ability to generate responses with a higher word count may have contributed to its high scientific score. However, long responses containing unnecessary details may cause confusion in patient education. Therefore, when evaluating the scientific depth of responses, attention was paid not only to the number of words but also to content density, evidence-based information presentation, and literature references. However, it should be noted that the relationship between response length and quality is not linear; unnecessarily lengthy and off-topic responses can have negative consequences in patient education [12].

In the accuracy category, it is concerning that ChatGPT-5 has a significantly lower average score compared to both Gemini-2.5 and DeepSeek-V3. ChatGPT-5's most noteworthy response was providing clear information that a fistula would mature and be ready for dialysis within four weeks in response to a question about fistula maturation (Question 8: What does it mean for a fistula to “mature”? How long does it take?). It did not provide a time frame and did not mention the valuable “Rules of 6” found in the guidelines [2]. This may indicate that ChatGPT-5 has a higher risk of producing hallucinations in this specific domain. Rater 3's evaluations were where this difference was most pronounced. One of the major criticisms of LLM use in medicine - the risk of hallucination - —has once again been confirmed by the results of our study [13, 14]. The statistically similar performance of Gemini-2.5 and DeepSeek-V3 in terms of accuracy suggests that these models may be more consistent in providing reliable information.

In contrast, Gemini-2.5 demonstrated a more balanced and moderate performance profile across the evaluated domains. While it did not exhibit pronounced strengths comparable to ChatGPT-5 in clarity or to DeepSeek-V3 in scientific depth, it also did not show marked weaknesses. This intermediate performance may suggest a more conservative response generation approach, prioritizing consistency over specialization, which could be advantageous in certain patient education contexts.

Interestingly, all three models scored similarly high in the clarity category, with no statistically significant difference observed between them. This finding demonstrates that current LLMs have generally reached a mature level of ability to express complex medical concepts in understandable language appropriate for the patient level. This is a promising result that strengthens their potential for use as a supportive tool, particularly for patient groups with low health literacy [15, 16].

The scoring variance among evaluators (e.g., Rater 1 not observing a significant difference in scientific accuracy, while other raters did) highlights the role of subjective judgment in evaluating medical information. Different clinical experiences and emphases can lead to different interpretations of the quality of a response. This highlights the value of study designs using multi-rater, blinded methods in evaluating LLM outputs and points to the limitations of assessments based on a single expert opinion.

Strengths and Limitations

This study’s main strength lies in its direct, blinded comparison of three LLMs in the context of hemodialysis vascular access, a field where patient education is critically important. The use of clinically relevant, patient-centered questions and independent evaluation by multiple expert raters enhances the reliability of the findings. Moreover, the separate assessment of accuracy, clarity, and scientific depth

allows for a clear delineation of each model's strengths and limitations in patient education. Nevertheless, this study has some limitations. First, the models' performance was evaluated at a specific point in time (August 2025). Since LLMs are constantly updated and model parameters (e.g., temperature settings) can affect the randomness of outputs, their performance may vary over time. Additionally, differences in model architecture, training datasets, and the rapid update cycles of LLMs may limit direct quantitative comparisons with previously published studies evaluating earlier model versions. Second, although the 25-question set used in the study is comprehensive, it may not cover all possible scenarios and complications related to hemodialysis access. Thirdly, the evaluator panel consisted solely of cardiovascular surgeons, and the exclusion of nephrologists and dialysis nurses is a factor that limits the generalizability of the study. Another limitation of this study is that intraclass correlation coefficient (ICC) analyses did not yield interpretable results due to the inherently subjective nature of expert-based evaluations of LLM-generated responses, which may have contributed to variability in the findings. Finally, the real-world impact of the responses on patient groups (comprehension, compliance, satisfaction) has not been measured; this is an important area for future research.

CONCLUSION

In conclusion, this study demonstrates that DeepSeek-V3 may be a more reliable source in terms of scientific depth and overall quality regarding hemodialysis vascular access; while ChatGPT-5's relative weakness in accuracy reinforces the fact that LLMs should be critically verified rather than automatically accepted for medical use. The success of all models in terms of clarity, however, shows that this technology holds promise as a support tool in patient education. However, ethical concerns must also be considered when using LLMs in patient education. Patients may place a high level of trust in AI-based information, which could create potential risks if incorrect or incomplete information is provided. Therefore, clinicians should review LLM-based educational materials and ensure the accuracy and clinical appropriateness of the information provided.

Ethics Approval and Consent to Participate

As this study was based solely on the evaluation of AI-generated responses and did not involve any patient data, ethical committee approval was not required.

Data Availability

All data generated or analyzed during this study are included in this published article. The data that support the findings of this study are available on request from the corresponding author, upon reasonable request.

Authors' Contribution

Study Conception: MHE; Study Design: MHE, ÖFR; Supervision: MHE, ÖFR; Funding: N/A; Materials: N/A; Data Collection and/or Processing: ÖFR, AG, FA, EB; Statistical Analysis and/or Data Interpretation: ÖFR; Literature Review: AG, FA, EB; Manuscript Preparation: MHE; and Critical Review: ÖFR, AG, FA, EB.

Conflict of Interest

The author(s) disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The author(s) disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgments

The authors have no acknowledgments to declare.

Generative Artificial Intelligence Statement

The large language model-based chatbots ChatGPT-5 (OpenAI, Microsoft Corporation, San Francisco, CA, USA, 2025), Gemini-2.5 (Google, Mountain View, CA, USA, 2025) and DeepSeek-V3 (DeepSeek, Beijing, CHINA, 2025) were used solely to respond to patient questions about hemodialysis vascular access. The final content was critically reviewed and approved by human authors, who take full responsibility for the integrity and accuracy of the work. The language model was not involved in authorship or editorial decision-making.

The all content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

Editor's Note

All statements made in this article are solely those of the authors and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Jager KJ, Kovesdy C, Langham R, Rosenberg M, Jha V, Zoccali C. A single number for advocacy and communication-worldwide more than 850 million individuals have kidney diseases. Kidney Int. 2019;96(5):1048-1050. doi: 10.1016/j.kint.2019.07.012.
2. Lok CE, Huber TS, Lee T, et al.; National Kidney Foundation. KDOQI Clinical Practice Guideline for Vascular Access: 2019 Update. Am J Kidney Dis. 2020 Apr;75(4 Suppl 2):S1-S164. doi: 10.1053/j.ajkd.2019.12.001.
3. Schmidli J, Widmer MK, Basile C, et al., ESVS Guidelines Committee, EEVS Guidelines Reviewers. Editor's Choice - Vascular Access: 2018 Clinical Practice Guidelines of the European Society for Vascular Surgery (ESVS). Eur J Vasc Endovasc Surg. 2018;55(6):757-818. doi: 10.1016/j.ejvs.2018.02.001.
4. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J Health Behav. 2007;31 Suppl 1:19-26. doi: 10.5555/ajhb.2007.31.supp.S19.
5. Skoumalova I, Madarasova Geckova A, Rosenberger J, et al. Health Literacy and Change in Health-Related Quality of Life in Dialysed Patients. Int J Environ Res Public Health. 2022;19(2):620. doi: 10.3390/ijerph19020620.
6. Zavacka M, Skoumalova I, Geckova AM, et al. Does Health Literacy of Hemodialyzed Patients Predict the Type of Their Vascular Access? A Cross-Sectional Study on Slovak Hemodialyzed Population. Int J Environ Res Public Health. 2020;17(2):675. doi: 10.3390/ijerph17020675.
7. Singh P, Goyal L, Mallick DC, et al. Artificial Intelligence in Nephrology: Clinical Applications and Challenges. Kidney Med. 2024;7(1):100927. doi: 10.1016/j.xkme.2024.100927.
8. Ji Z, Lee N, Frieske R, et al. Survey of Hallucination in Natural Language Generation. ACM Comput Surv. 2023;55(12):248. doi: 10.1145/3571730.
9. Yeo YH, Samaan JS, Ng WH, et al. Assessing the performance of ChatGPT in answering questions regarding cirrhosis and hepatocellular carcinoma. Clin Mol Hepatol. 2023;29(3):721-732. doi: 10.3350/cmh.2023.0089.
10. Akçay O, Öztürk Ö, Acar T, Gürsoy S. Accuracy and Reliability of ChatGPT in Answering Patient Questions About Lung Cancer and Its Surgery: An Expert Panel Evaluation by Thoracic Surgeons. J Cancer Educ. 2025 Jul 4. doi: 10.1007/s13187-025-02682-3.
11. Almagazzachi A, Mustafa A, Eighaei Sedeh A, et al. Generative Artificial Intelligence in Patient Education: ChatGPT Takes on Hypertension Questions. Cureus. 2024;16(2):e53441. doi: 10.7759/cureus.53441.
12. Sivaramakrishnan G, Almuqahwi M, Ansari S, Lubbad M, Alagamawy E, Sridharan K. Assessing the power of AI: a comparative evaluation of large language models in generating patient education materials in dentistry. BDJ Open. 2025;11(1):59. doi: 10.1038/s41405-025-00349-1.
13. Sallam M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. Healthcare (Basel). 2023;11(6):887. doi: 10.3390/healthcare11060887.
14. Huang Y, Gomaa A, Semrau S, et al. Benchmarking ChatGPT-4 on a radiation oncology in-training exam and Red Journal Gray Zone cases: potentials and challenges for ai-assisted medical education and decision making in radiation oncology. Front Oncol. 2023;13:1265024. doi: 10.3389/fonc.2023.1265024.
15. Swisher AR, Wu AW, Liu GC, Lee MK, Carle TR, Tang DM. Enhancing Health Literacy: Evaluating the Readability of Patient Handouts Revised by ChatGPT's Large Language Model. Otolaryngol Head Neck Surg. 2024;171(6):1751-1757. doi: 10.1002/ohn.927.
16. Eid K, Eid A, Wang D, Raiker RS, Chen S, Nguyen J. Optimizing Ophthalmology Patient Education via ChatBot-Generated Materials: Readability Analysis of AI-Generated Patient Education Materials and The American Society of Ophthalmic Plastic and Reconstructive Surgery Patient Brochures. Ophthalmic Plast Reconstr Surg. 2024;40(2):212-216. doi: 10.1097/IOP.0000000000002549.

The European Research Journal 2026;12(7):739-751 Research Article

DOI: https://doi.org/10.18621/eurj.1815977 *Endocrinology and Metabolism*

Atherogenic Burden and Insulin Resistance in Non-Obese Women with Polycystic Ovary Syndrome: A Comparative Study with Healthy Controls

Burak Andaç[1], Mehtap Navdar Başaran[2], Gözde Nur Eren[3]

[1]Department of Endocrinology and Metabolism, Ordu University Training and Research Hospital, Ordu, Türkiye; [2]Department of Endocrinology and Metabolism, Ordu Medical Park Hospital, Ordu, Türkiye; [3]Department of Internal Medicine, Ordu University Training and Research Hospital, Ordu, Türkiye

ABSTRACT

Objective: Polycystic ovary syndrome (PCOS) is associated with elevated cardiovascular disease (CVD) risk due to metabolic derangements. Although obesity contributes to CVD, the independent contribution of PCOS remains controversial. This study aimed to compare insulin resistance (IR) and atherogenic lipid indices across body mass index (BMI) categories within PCOS, to evaluate the diagnostic performance of atherogenic indices between non-obese PCOS and healthy controls, and to derive exploratory, data-driven thresholds that may inform cardiovascular risk assessment in non-obese PCOS.

Methods: This single-center case-control study enrolled 65 treatment-naïve women newly diagnosed with PCOS (Rotterdam criteria 2003) and 100 age-matched healthy controls. A panel of IR and composite atherogenic lipid indices (e.g., Lipid Accumulation Product [LAP], Visceral Adiposity Index [VAI], Triglyceride-Glucose [TyG] Index) was compared between groups.

Results: Within PCOS, atherogenic lipid indices were similar in obese versus non-obese subgroups. Non-obese PCOS demonstrated significantly higher atherogenic lipid and IR indices compared to controls (all $P<0.05$). Receiver operating characteristic (ROC) curve analysis identified LAP (Area under the curve [AUC] =0.747, 95% confidence intervals [CI]: 0.647-0.844, threshold=18.05), VAI (AUC=0.707, 95% CI: 0.605-0.811, threshold=1.15), and TyG (AUC=0.701, 95% CI: 0.583-0.814, threshold=8.29) as the three best indices distinguishing non-obese PCOS from healthy controls.

Conclusion: Elevated atherogenic lipid and IR indices in non-obese PCOS compared to non-obese controls, alongside similar atherogenic markers between obese and non-obese PCOS groups are consistent with the notion that PCOS may independently contribute to atherogenesis beyond obesity-related mechanisms. In this exploratory analysis, data-driven thresholds for LAP, VAI and TyG were identified in non-obese PCOS. These values require external validation in larger, multi-center cohorts before they can be considered for routine CVD risk screening or monitoring.

Keywords: Polycystic Ovary Syndrome, Lipid Accumulation Product, Triglyceride-Glucose Index, Visceral Adiposity Index

Submitted: November 5, 2025 **Accepted:** December 7, 2025 **Published Online:** December 29, 2025

How to cite this article: Andaç B, Navdar Başaran M, Eren GN. Atherogenic Burden and Insulin Resistance in Non-Obese Women with Polycystic Ovary Syndrome: A Comparative Study with Healthy Controls. *Eur Res J.* 2026;12(7):739-751. doi: 10.18621/eurj.1815977

Corresponding author: Burak Andaç, MD., Phone: +90 452 225 01 85, E-mail: drburakandac87@gmail.com

Available Online at https://www.eurj.org.tr

Polycystic ovary syndrome (PCOS) is a prevalent, multifaceted endocrine disorder in women of reproductive age, with global estimates ranging from 4% to 21% depending on the diagnostic criteria applied and the characteristics of the studied population [1]. PCOS exhibits marked heterogeneity, encompassing a broad range of clinical manifestations and diverse phenotypes. Clinically, it is characterized by oligomenorrhea and hyperandrogenism, alongside cardiovascular risk factors including glucose intolerance, insulin resistance (IR), obesity, dyslipidemia, metabolic dysfunction-associated steatotic liver disease, and obstructive sleep apnea [1, 2]. Despite the absence of a controlled, systematic study to delineate the exact prevalence, the prevailing literature indicates that ≥50% of women with PCOS are obese. Furthermore, relative to women without PCOS, the majority exhibit hyperinsulinemia and IR irrespective of obesity [3, 4]. A meta-analysis has confirmed elevated risks of coronary heart disease (CHD) and stroke in PCOS populations [4]. Factors such as obesity, IR, type 2 diabetes mellitus (DM), and dyslipidemia appear to contribute to the heightened susceptibility of this population to CHD [3, 5]. However, the independent contribution of PCOS to cardiovascular risk, distinct from the confounding influence of obesity, remains incompletely understood. Multiple studies demonstrated that non-obese women with PCOS exhibit higher levels of cardiovascular risk markers - including inflammatory markers, endothelial dysfunction, and subclinical atherosclerosis - than their non-PCOS counterparts, independent of body mass index (BMI) [6-8].

Anthropometric (e.g., BMI, waist circumference [WC]), atherogenic (e.g., Castelli risk index, Atherogenic Index of Plasma [AIP], Atherogenic Coefficient [AC]), and IR indices (e.g., Homeostatic Model Assessment of Insulin Resistance [HOMA-IR], Triglyceride-Glucose [TyG] index, Metabolic Score for Insulin Resistance [METS-IR], Visceral Adiposity Index [VAI], Lipid Accumulation Product [LAP]) are increasingly used as non-invasive, cost-effective tools for predicting cardiovascular disease (CVD) risk across diverse populations. Recent large-scale cohort studies demonstrate that these indices are independently associated with incident CVD and mortality [9-11]. Furthermore, these indices have been shown to reflect metabolic disturbances in PCOS [12-15]. However, comprehensive evaluations comparing multiple validated atherogenic and IR indices across BMI categories within PCOS cohorts remain scarce. Moreover, defining clinically applicable diagnostic cut-offs for these indices specifically in non-obese PCOS would facilitate early risk stratification and guide targeted preventive interventions.

The present study aimed to address these knowledge gaps through four primary objectives: (1) to comprehensively compare markers of IR/atherogenic indices in PCOS and healthy controls (including obese and non-obese), (2) to compare these markers between obese and non-obese subgroups within the PCOS cohort, (3) to evaluate the diagnostic performance of the indices in discriminating between non-obese PCOS and healthy controls, and (4) to derive exploratory, data-driven threshold values that may help inform clinical screening and longitudinal monitoring of cardiovascular risk in non-obese PCOS populations.

METHODS

Study Design and Participants

This single-center case-control study was conducted using data from patients examined between December 2023 and July 2025. The study population comprised 65 treatment-naive women newly diagnosed with PCOS and 100 age-matched healthy controls. PCOS diagnosis was established according to the Rotterdam criteria 2003, requiring the presence of at least two of three features: (1) oligo-ovulation or anovulation, (2) clinical and/or biochemical hyperandrogenism, and (3) polycystic ovaries on ultrasonography, after exclusion of other etiologies [16]. All PCOS participants were treatment-naive at enrollment to eliminate the confounding effects of medications on metabolic parameters.

Inclusion criteria for the PCOS group included: age 18-45 years, fulfillment of Rotterdam diagnostic criteria, absence of any medical treatment for PCOS or metabolic disorders in the preceding six months. Exclusion criteria encompassed: pregnancy, lactation, use of lipid-lowering drugs, strenuous exercise habits, history of previous or current oral contraceptive use, patients using glucocorticoids and drugs that may

affect glucose metabolism, decompensated thyroid disease, patients with other causes of androgen excess (classic or nonclassic congenital adrenal hyperplasia, hyperprolactinemia, Cushing's syndrome, androgen-secreting tumors, idiopathic hirsutism), patients diagnosed with depression, and patients using medications with a diagnosis of depression, chronic inflammatory diseases, hepatic or renal dysfunction, and smoking or excessive alcohol consumption.

Healthy controls were recruited from hospital staff and community volunteers undergoing routine health examinations. Control inclusion criteria included: age 18-45 years, regular menstrual cycles (26-35 days), absence of clinical or biochemical hyperandrogenism, and absence of any chronic medical conditions or regular medication use. The same exclusion criteria applied to the PCOS group were also applied to the control group.

Anthropometric Measurements

BMI was calculated as weight (kg) divided by height squared (m²). Obesity was defined as BMI ≥30 kg/m² and non-obesity as BMI <30 kg/m², in accordance with commonly used World Health Organization (WHO) criteria. Within the PCOS group, women were categorised as obese (BMI ≥30 kg/m²) or non-obese (BMI <30 kg/m²) for subgroup analyses. Obesity subclasses were not analysed separately because of the limited number of patients in each category. WC was measured at the midpoint between the lowest rib and iliac crest during mid-expiration using a non-stretchable tape measure to the nearest 1 cm. The waist-to-height ratio (WHtR) was computed as waist circumference (cm) divided by height (cm).

Biochemical Analyses

Blood samples were collected after a minimum 12-hour overnight fast. Glycated hemoglobin (HbA1c) was quantified on a Tosoh G8 high-performance liquid chromatography (HPLC) analyzer utilizing manufacturer-supplied reagents. Serum glucose, total cholesterol (TC), low density lipoprotein (LDL)-cholesterol (LDL-C), high density lipoprotein (HDL)-cholesterol (HDL-C), triglycerides (TG), fasting plasma glucose (FPG) were assayed on a Roche C702 clinical chemistry platform using the corresponding proprietary kits. Serum insulin concentrations were determined by chemiluminescent immunoassay with analytical sensitivity of 0.5 µIU/mL.

Calculation of Insulin Resistance and Atherogenic Indices

All laboratory values are reported in mg/dL; for indices whose original definitions required SI units, TG and HDL were converted to mmol/L only for index calculations, using TG [mmol/L] = TG [mg/dL] × 0.01129 and HDL [mmol/L] = HDL [mg/dL] × 0.02586. All indices are unitless; TG/HDL were internally converted to mmol/L only for AIP, VAI, and LAP as per original definitions."

Insulin Resistance Indices

-HOMA-IR = (FPG [mg/dL] × fasting insulin [µIU/mL]) / 405 [17]

-Quantitative Insulin Sensitivity Check Index (QUICKI) = 1 / (log(fasting insulin [µIU/mL]) + log(FPG [mg/dL])) [18]

-Fasting Glucose–Insulin Ratio (FG-IR) = FPG (mg/dL) / fasting insulin (µIU/mL) [12]

-Metabolic Score for Insulin Resistance (METS-IR) = (Ln[2 × FPG (mg/dL) + fasting insulin (µIU/mL)] × TG (mg/dL)) / Ln(HDL-C [mg/dL]) [11]

Atherogenic Lipid Indices

-TyG index = Ln[TG (mg/dL) × FPG (mg/dL) / 2] [9]

-TyG-BMI = TyG index × BMI (kg/m²) [12]

-TG/HDL ratio = TG (mg/dL) / HDL-C (mg/dL) [12]

-Atherogenic Index of Plasma AIP = Log_{10}[TG (mmol/L) / HDL-C (mmol/L)] [12]

-LAP (women) = [WC (cm) − 58] × TG (mmol/L) [10]

-Visceral Adiposity Index (VAI) (women) = [WC (cm) / (36.58 + 1.89 × BMI)] × [TG (mmol/L) / 0.81] × [1.52 / HDL-C (mmol/L)] [10]

-Castelli Risk Index I = TC (mg/dL) / HDL-C (mg/dL) [19]

-Castelli Risk Index II = LDL-C (mg/dL) / HDL-C (mg/dL) [19]

-AC = (TC (mg/dL) − HDL-C (mg/dL)) / HDL-C (mg/dL) [12]

-Lipoprotein Combine Index (LCI) = TC (mg/dL) × TG (mg/dL) × LDL-C (mg/dL) / HDL-C (mg/dL) [12]

-non−HDL−C = TC(mg/dL) − HDL−C (mg/dL) [12]

In this study, the term 'atherogenic burden' is used

in a descriptive manner to denote the overall pattern of classical lipid parameters (TC, LDL-C, HDL-C, TG, non-HDL-C) and derived atherogenic indices (TyG, TG/HDL-C, AIP, Castelli risk index I and II, AC, and LCI) that have been associated with increased cardiovascular risk.

Statistical Analysis

Normality of continuous variables was assessed using the Kolmogorov-Smirnov test and visual inspection of Q-Q plots. Data are presented as mean±standard deviation for normally distributed variables, median (interquartile range) for non-normally distributed variables, and frequencies with percentages for categorical variables. Comparisons between two independent groups were performed using Student's t-test for normally distributed variables, Welch's t-test when variances were unequal, and Mann-Whitney U test for non-normally distributed variables. Categorical variables were compared using chi-square or Fisher's exact test as appropriate. Effect sizes were reported for all between-group comparisons. For parametric contrasts (Student/Welch t-test), we calculated Hedges' g

TABLE 1. Comparisons of Anthropometric Measurements and Biochemical Parameters in PCOS and Control Groups

Variables	PCOS group (n=65)	Control group (n=100)	P-value	Effect size (95% CI)
Age (years)	25.00 [22.00–29.00]	26.00 [18.00–37.00]	0.145[a]	**$r_{(rb)} \approx -0.13$** (−0.32, 0.05)
BMI (kg/m^2)	28.23 [24.35–33.46]	24.16 [21.31–28.36]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
WHtR	0.59 [0.53–0.67]	0.49 [0.45–0.56]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
WC (cm)	95.00 [88.00–104.00]	80.00 [73.75–90.00]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
Fasting glucose (mg/dL)	90.00 [85.00–99.00]	90.00 [85.00–94.00]	0.175[a]	**$r_{(rb)} \approx 0.13$** (−0.06, 0.31)
Fasting insulin (mU/L)	13.60 [9.92–23.20]	8.91 [6.72–12.05]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
HbA1c (%)	5.30 [5.10–5.50]	5.30 [5.10–5.60]	0.607[a]	**$r_{(rb)} \approx 0.05$** (−0.13, 0.23)
TC (mg/dL)	180.00 [158.00–205.00]	172.50 [157.75–196.00]	0.340[a]	**$r_{(rb)} \approx 0.09$** (−0.09, 0.27)
HDL-C (mg/dL)	50.65±9.47	55.91±13.13	**0.003[b]**	**Hedges g = −0.44** (−0.76, −0.13)
LDL-C (mg/dL)	104.00 [87.00–125.00]	100.50 [84.75–117.50]	0.361[a]	**$r_{(rb)} \approx 0.08$** (−0.10, 0.26)
Triglyceride (mg/dL)	101.00 [70.00–138.00]	77.50 [56.00–96.12]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
non-HDL-C(mg/dL)	132.00 [102.00–153.00]	115.00 [100.00–140.25]	0.085[a]	**$r_{(rb)} \approx 0.16$** (−0.02, 0.34)

Data are shown as mean±standard deviation or median [IQR] where appropriate. BMI, body mass index; HbA1c, glycated hemoglobin; HDL-C, high-density lipoprotein cholesterol; LDL-C, low-density lipoprotein cholesterol; PCOS, polycystic ovary syndrome; TC, total cholesterol; WC, waist circumference; WHtR, Waist/height ratio. Distribution was evaluated by KS test and variables were reported accordingly. [a]Mann–Whitney U, [b]Welch t-test. Statistically significant P-values are shown in bold.

TABLE 2. Comparisons of Atherogenic/Insulin Resistance Indices in PCOS and Control Groups

Variables	PCOS group (n=65)	Control group (n=100)	P-value	Effect size (95% CI)
TyG index	8.40 [8.05–8.83]	8.14 [7.79–8.41]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
TyG-BMI	240.39 [201.15–283.89]	193.20 [171.79–240.99]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
TG/HDL ratio	2.04 [1.25–3.07]	1.35 [0.96–1.97]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
AIP (log10, mmol/L)	-0.05 [-0.26–0.13]	-0.23 [-0.38–-0.06]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
METS-IR	41.62 [34.60–48.80]	32.39 [29.38–40.77]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
LAP	37.32 [23.95–66.99]	17.21 [10.77–28.80]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
VAI	1.64 [1.03–2.51]	1.06 [0.76–1.59]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
HOMA-IR	3.15 [2.06–5.16]	1.90 [1.41–2.75]	**<0.001[a]**	**$r_{(rb)} \geq 0.30$** (≈0.12, 0.49)
QUICKI	0.32±0.04	0.35±0.03	**<0.001[b]**	**Hedges g = −0.87** (−1.20, −0.54)
FG-IR	6.91 [3.84–9.44]	9.69 [7.61–13.20]	**<0.001[a]**	**$r_{(rb)} \leq -0.30$** (≈−0.49, −0.12)
Castelli I	3.58 [2.83–4.35]	3.07 [2.63–3.87]	**0.013[a]**	**$r_{(rb)} \approx 0.23$** (0.05, 0.41)
Castelli II	2.08 [1.62–2.76]	1.81 [1.36–2.47]	**0.040[a]**	**$r_{(rb)} \approx 0.19$** (0.01, 0.37)
AC	2.58 [1.83–3.35]	2.07 [1.63–2.87]	**0.013[a]**	**$r_{(rb)} \approx 0.23$** (0.05, 0.41)
LCI	43937.58 [17758.93–70031.71]	22694.63 [12864.99–44917.91]	**0.004[a]**	**$r_{(rb)} \approx 0.16$** (−0.02, 0.34)

Data are shown as mean±standard deviation or median [IQR] where appropriate. AC, atherogenic coefficient; AIP, atherogenic index; FG-IR, fasting glucose–insulin ratio; HDL, high-density lipoprotein; HOMA-IR, homeostatic model assessment for insulin resistance; LAP, lipid accumulation product; LCI, lipid combination index; METS-IR, Metabolic score for insulin resistance; QUICKI, quantitative insulin sensitivity check index; TG, triglyceride; TyG, triglyceride-glucose; PCOS, polycystic ovary syndrome; VAI, Visceral adiposity index. Distribution was evaluated by KS test and variables were reported accordingly.

METS-IR = (Ln[2 × FPG + fasting insulin (μIU/mL)] × TG / Ln(HDL-C); AIP = Log_{10}[TG / HDL-C]; LAP= [WC − 58] × TG; VAI= [WC / (36.58 + 1.89 × BMI)] × [TG / 0.81] × [1.52 / HDL-C]; Castelli Risk Index I = TC / HDL-C; Castelli Risk Index II = LDL-C / HDL-C; AC = (TC − HDL-C) / HDL-C; LCI = TC × TG × LDL-C / HDL-C.

[a]Mann–Whitney U, [b]Student t-test. Statistically significant P-values are shown in bold.

(bias-corrected Cohen's d). For non-parametric contrasts (Mann–Whitney U), we calculated the rank-biserial correlation. Two-sided 95% confidence intervals were derived via bootstrap resampling (5,000 iterations). Effect sizes were oriented such that positive values indicate higher levels in the first-listed group for each comparison (PCOS vs controls in Tables 1A–B; BMI <30 vs BMI ≥30 kg/m² in Table 2; non-obese PCOS vs non-obese controls in Table 3). Thresholds for magnitude were |g|≈0.2/0.5/0.8 and |r_{rb}|≈0.1/0.3/0.5 (small/medium/large). Receiver operating characteristic (ROC) curve analysis was conducted to evaluate diagnostic performance of atherogenic indices in discriminating non-obese PCOS from healthy controls. Area under the curve (AUC) with 95% confidence intervals (CI) was calculated. Optimal cutoff values were determined using Youden's index (sensitivity + specificity − 1), maximizing both sensitivity and specificity. DeLong's test was used to compare AUCs between different indices. Because LAP exhibited the highest AUC among all indices, it was used as the reference curve in DeLong comparisons. Statistical analyses were performed using SPSS version 22.0 (IBM Corp., Armonk, NY, USA) and P-values <0.05 were considered statistically significant.

RESULTS

In a comparative analysis between women diagnosed with PCOS (n=65) and a cohort of healthy controls (n=100), the PCOS group had significantly higher values for key indicators of atherogenic dyslipidemia and IR. Detailed comparisons of anthropometric measurements, biochemical parameters, and the atherogenic/IR indices are summarized in Tables 1 and 2.

Upon dividing the PCOS group into obese and non-obese categories based on BMI, atherogenic lipid indices did not differ significantly between groups (Specifically, TyG, TG/HDL, AIP, Castelli I, Castelli II, AC, LCI, and non-HDL; P=0.295, P=0.089, P=0.085, P=0.127, P=0.287, P=0.127, P=0.504, and P=0.842, respectively) (Table 3).

Non-obese women with PCOS (n=37) were then compared with non-obese controls (n=81). In non-obese women, TG, TC and non-HDL-C were significantly higher in the PCOS group than in BMI-matched controls, whereas LDL-C did not differ significantly between groups (Table 3). Non-obese women with PCOS also had significantly higher values for almost all atherogenic and IR indices evaluated. Specifically, TyG, TG/HDL, AIP, METS-IR, LAP, VAI, HOMA-IR, Castelli I–II, AC, non-HDL, and LCI were higher in the PCOS cohort (P=<0.001, P=0.002, P=0.002, P=0.017, P<0.001, P<0.001, P<0.001, P=0.016, P=0.032, P=0.016, and P=0.018, respectively) (Table 4).

ROC curve analysis with AUC estimation was performed to assess the ability of atherogenic indices to discriminate non-obese PCOS from non-obese controls and to determine optimal cut-off values. The top three indices with the highest AUC values were: LAP (AUC = 0.747, 95% CI: 0.647–0.844; optimal threshold = 18.05; sensitivity = 78.4%; specificity = 63.0%), VAI (AUC = 0.707, 95% CI: 0.605–0.811; threshold = 1.15; sensitivity = 64.9%; specificity = 66.7%), and TyG (AUC = 0.701, 95% CI: 0.583–0.814; threshold = 8.29; sensitivity = 59.5%; specificity = 79.0%) (Table 5, Figure 1).

DISCUSSION

The present study provides evidence that non-obese women with PCOS exhibit significantly elevated atherogenic burden and insulin resistance compared to healthy controls, while atherogenic lipid indices remain comparably elevated across obesity categories within PCOS. These findings are consistent with the hypothesis that PCOS may independently contribute to cardiovascular risk beyond obesity-related mechanisms.

PCOS is strongly associated with obesity and with the clinical sequelae of IR. IR contributes to disordered steroidogenesis and consequent androgen excess [1, 3]. Obesity-independent cardiovascular risk in PCOS may stem from several underlying mechanisms, including IR, hyperandrogenism, chronic low-grade inflammation, and endothelial dysfunction [20-26]. PCOS is linked to a more atherogenic lipid profile (low HDL-C and high TG concentrations) and higher blood pressure, even in the absence of obesity, further increasing cardiovascular risk [1, 3]. Moreover, relative to BMI- and IR–matched women without PCOS, women with PCOS

TABLE 3. Comparison of Atherogenic / Insulin Resistance Indices According to the Presence of Obesity in the PCOS Patient Group

Variables	BMI<30 kg/m² (n=37)	BMI≥30 kg/m² (n=28)	P-value	Effect size (95% CI)
HOMA-IR	2.97 [1.75–3.89]	3.67 [2.49–6.93]	**0.0381**[a]	**$r_{(rb)} \approx -0.10$** (−0.19, −0.01)
QUICKI	0.33±0.04	0.31±0.03	**0.0270**[b]	**Hedges g = +0.55** (0.05, 1.05)
FG-IR	7.67 [5.25–9.72]	4.78 [3.43–8.13]	**0.0215**[a]	**$r_{(rb)} \approx +0.11$** (0.02, 0.19)
METS-IR	34.88 [30.61–38.46]	50.65 [46.73–57.28]	**<0.0001**[a]	**$r_{(rb)} \approx -0.18$** (−0.27, −0.09)
TyG	8.36±0.51	8.50±0.53	0.2951[b]	Hedges g = −0.27 (−0.76, 0.23)
TyG-BMI	205.33±31.64	298.98±39.98	**<0.0001**[b]	**Hedges g = −2.61** (−3.28, −1.94)
VAI	1.38 [1.02–2.19]	2.06 [1.14–3.01]	0.0815[a]	$r_{(rb)} \approx -0.08$ (−0.17, 0.00)
LAP	26.58 [18.63–47.19]	64.77 [36.94–85.49]	**<0.0001**[a]	**$r_{(rb)} \approx -0.18$** (−0.27, −0.09)
TG/HDL-C	1.73 [1.17–2.63]	2.35 [1.33–3.52]	0.0887[a]	$r_{(rb)} \approx -0.08$ (−0.16, 0.00)
AIP (log10, mmol/L)	-0.11±0.24	-0.00±0.27	0.0851[b]	Hedges g = −0.43 (−0.93, 0.07)
Castelli I	3.51±0.81	3.86±1.05	0.1274[b]	Hedges g = −0.37 (−0.86, 0.12)
Castelli II	2.09±0.63	2.29±0.82	0.2872[b]	Hedges g = −0.28 (−0.77, 0.22)
AC	2.51±0.81	2.86±1.05	0.1274[b]	Hedges g = −0.37 (−0.86, 0.12)
LCI	37847.20 [19183.50–62136.36]	55412.44 [16306.23–73022.92]	0.5035[a]	$r_{(rb)} \approx -0.03$ (−0.12, 0.00)
non-HDL (mg/dL)	129.51±31.12	127.86±35.48	0.8421[b]	Hedges g = +0.05 (−0.44, 0.54)

Data are shown as mean±standard deviation or median [IQR] where appropriate. AC, atherogenic coefficient; AIP, atherogenic index; BMI, body mass index; FG-IR, fasting glucose–insulin ratio; HDL-C, high-density lipoprotein cholesterol; HOMA-IR, homeostatic model assessment for insulin resistance; LAP, lipid accumulation product; LCI, lipid combination index; METS-IR, Metabolic score for insulin resistance; QUICKI, quantitative insulin sensitivity check index; TG, triglyceride; TyG, triglyceride-glucose; PCOS, polycystic ovary syndrome; VAI, Visceral adiposity index.

Distribution was evaluated by KS test and variables were reported accordingly.

METS-IR = (Ln[2 × FPG + fasting insulin (μIU/mL)] × TG / Ln(HDL-C); AIP = Log_{10}[TG / HDL-C]; LAP= [WC − 58] × TG; VAI= [WC / (36.58 + 1.89 × BMI)] × [TG / 0.81] × [1.52 / HDL-C]; Castelli Risk Index I = TC / HDL-C; Castelli Risk Index II = LDL-C / HDL-C; AC = (TC − HDL-C) / HDL-C; LCI = TC × TG × LDL-C / HDL-C.

[a]Mann–Whitney U, [b]Student t-test. Statistically significant P-values are shown in bold.

TABLE 4. Atherogenic/Insulin Resistance Indices and Classical Lipid Parameters in Non-Obese PCOS and Non-Obese Control Groups

Variables	Non-obese PCOS (n=37)	Non-obese control (n=81)	P-value	Effect size (95% CI)
Insulin resistance indices				
TyG	8.30 [8.05–8.69]	8.03 [7.76–8.26]	**<0.001**[a]	**$r_{(rb)} \approx +0.38$** (0.15, 0.60)
TyG-BMI	205.98 [181.53–226.42]	183.37 [170.27–204.98]	**0.004**[a]	**$r_{(rb)} \approx +0.33$** (0.11, 0.56)
METS-IR	34.88 [30.61–38.46]	31.64 [28.42–35.29]	**0.017**[a]	**$r_{(rb)} \approx +0.28$** (0.05, 0.50)
HOMA-IR	2.97 [1.75–3.89]	1.80 [1.33–2.31]	**<0.001**[a]	**$r_{(rb)} \approx +0.38$** (0.15, 0.60)
QUICKI	0.33±0.04	0.35±0.03	**0.002**[b]	**Hedges g = −0.59** (−0.99, −0.20)
FG-IR	7.67 [5.25–9.72]	10.00 [8.42–14.02]	**<0.001**[a]	**$r_{(rb)} \approx -0.38$** (−0.60, −0.15)
Atherogenic indices				
TG/HDL-C ratio	1.73 [1.17–2.63]	1.19 [0.93–1.71]	**0.002**[a]	**$r_{(rb)} \approx +0.36$** (0.13, 0.58)
AIP (log10, mmol/L)	-0.12 [-0.29–0.06]	-0.28 [-0.39–-0.13]	**0.002**[a]	**$r_{(rb)} \approx +0.36$** (0.13, 0.58)
LAP	26.58 [18.63–47.19]	14.63 [9.14–23.48]	**<0.001**[a]	**$r_{(rb)} \approx +0.38$** (0.15, 0.60)
VAI	1.38 [1.02–2.19]	0.96 [0.73–1.43]	**<0.001**[a]	**$r_{(rb)} \approx +0.38$** (0.15, 0.60)
Castelli I	3.39 [2.83–4.26]	2.94 [2.55–3.53]	**0.016**[a]	**$r_{(rb)} \approx +0.28$** (0.05, 0.50)
Castelli II	1.98 [1.64–2.73]	1.71 [1.33–2.13]	**0.032**[a]	**$r_{(rb)} \approx +0.25$** (0.02, 0.47)
AC	2.39 [1.83–3.26]	1.94 [1.55–2.53]	**0.016**[a]	**$r_{(rb)} \approx +0.28$** (0.05, 0.50)
LCI	37847.20 [19183.50–62136.36]	18507.27 [12010.58–34538.40]	**0.002**[a]	**$r_{(rb)} \approx +0.36$** (0.13, 0.58)
Classical lipid parameters				
HDL-C (mg/dL)	53.46±8.19	56.71±13.30	0.106[b]	Hedges g = −0.27 (−0.66, 0.12)
TG (mg/dL)	101.00 [69.00–123.00]	72.00 [55.00–83.00]	**<0.001**[a]	**$r_{(rb)} \approx +0.38$** (0.15, 0.60)
TC (mg/dL)	186.00 [166.00–205.00]	169.00 [155.00–186.00]	**0.037**[a]	**$r_{(rb)} \approx +0.24$** (0.02, 0.47)
LDL-C (mg/dL)	110.00 [88.00–127.00]	98.00 [83.00–113.00]	0.059[a]	$r_{(rb)} \approx +0.22$ (0.00, 0.44)
non-HDL-C (mg/dL)	132.00 [104.00–153.00]	111.00 [97.00–135.00]	**0.018**[a]	**$r_{(rb)} \approx +0.27$** (0.05, 0.50)

Data are shown as mean±standard deviation or median [IQR] where appropriate. AC, atherogenic coefficient; AIP, atherogenic index; BMI, body mass index; FG-IR, fasting glucose–insulin ratio; HDL-C, high-density lipoprotein cholesterol; HOMA-IR, homeostatic model assessment for insulin resistance; LAP, lipid accumulation product; LCI, lipid combination index; LDL-C, low-density lipoprotein cholesterol; METS-IR, Metabolic score for insulin resistance; QUICKI, quantitative insulin sensitivity check index; TC, total cholesterol; TG, triglyceride; TyG, triglyceride-glucose; PCOS, polycystic ovary syndrome; VAI, Visceral adiposity index. Distribution was evaluated by KS test and variables were reported accordingly.

METS-IR = (Ln[2 × FPG + fasting insulin (μIU/mL)] × TG / Ln(HDL-C); AIP = Log_{10}[TG / HDL-C]; LAP= [WC − 58] × TG; VAI= [WC / (36.58 + 1.89 × BMI)] × [TG / 0.81] × [1.52 / HDL-C]; Castelli Risk Index I = TC / HDL-C; Castelli Risk Index II = LDL-C / HDL-C; AC = (TC − HDL-C) / HDL-C; LCI = TC × TG × LDL-C / HDL-C.

[a]Mann-Whitney U, [b]Welch t-test. Statistically significant P-values are shown in bold.

TABLE 5. ROC Analysis for Threshold Values of Atherogenic/ınsulin Resistance Indices Between Non-Obese PCOS and Control Groups

Index	AUC	%95 CI Lower	%95 CI Upper	Youden Cut-off	Sensitivity	Specificity	ΔAUC vs LAP	P-value (DeLong, vs LAP)
LAP	0.747	0.647	0.844	18.053	78%	63%	–	–
VAI	0.707	0.605	0.811	1.150	65%	67%	0.040	0.584
TyG index	0.701	0.583	0.814	8.289	59%	79%	0.046	0.577
AIP	0.678	0.566	0.792	-0.055	46%	85%	0.068	0.429
TG/HDL-C	0.678	0.566	0.792	2.020	46%	85%	0.069	0.428
LCI	0.677	0.578	0.781	37 847	51%	80%	0.069	0.414
Castelli I	0.638	0.548	0.747	3.295	59%	69%	0.108	0.308
AC	0.638	0.548	0.747	2.295	59%	69%	0.108	0.308
METS-IR	0.638	0.514	0.736	32.979	70%	63%	0.109	0.203
Castelli II	0.623	0.530	0.733	1.571	81%	44%	0.123	0.275

AC, atherogenic coefficient; AIP, atherogenic index; AUC, area under the curve; BMI, body mass index; CI, confidence interval; HDL-C, high-density lipoprotein cholesterol; LAP, lipid accumulation product; LCI, lipid combination index; LDL-C, low-density lipoprotein cholesterol; METS-IR, Metabolic score for insulin resistance; TC, total cholesterol; TG, triglyceride; TyG, triglyceride-glucose; PCOS, polycystic ovary syndrome; ROC, receiver operating characteristic; VAI, Visceral adiposity index.

METS-IR = (Ln[2 × FPG + fasting insulin (μIU/mL)] × TG / Ln(HDL-C); AIP = Log_{10}[TG / HDL-C]; LAP= [WC − 58] × TG; VAI= [WC / (36.58 + 1.89 × BMI)] × [TG / 0.81] × [1.52 / HDL-C]; Castelli Risk Index I = TC / HDL-C; Castelli Risk Index II = LDL-C / HDL-C; AC = (TC − HDL-C) / HDL-C; LCI = TC × TG × LDL-C / HDL-C.

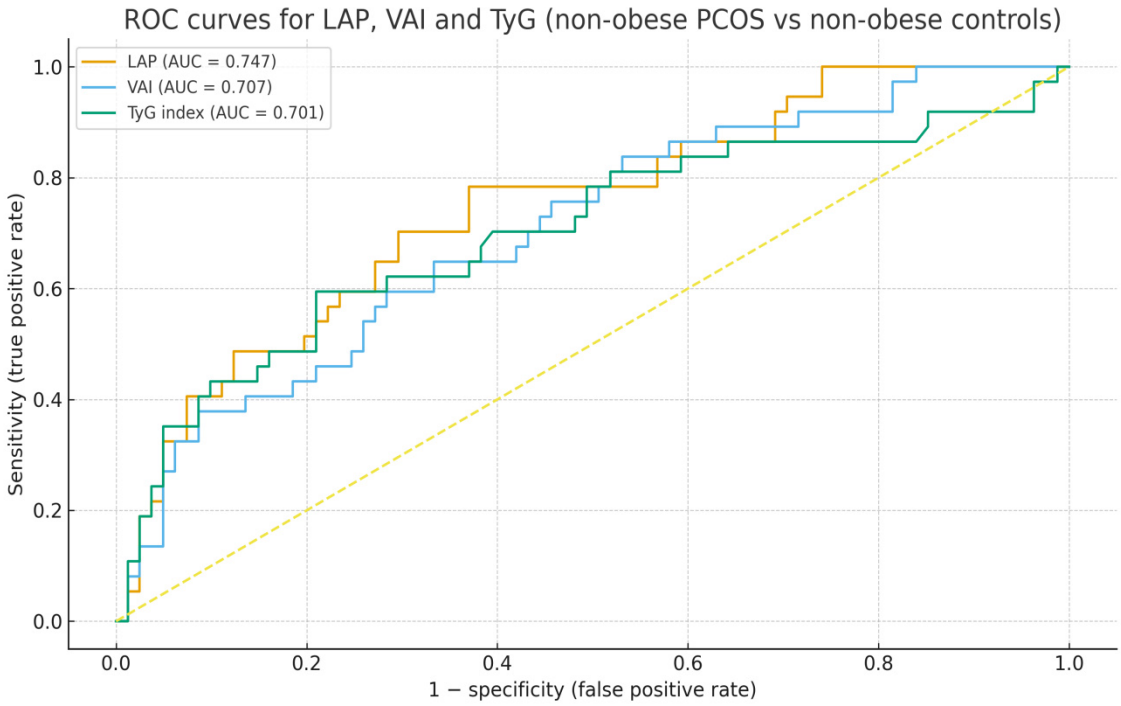

FIGURE 1. ROC analysis for threshold values of atherogenic lipid/insulin resistance indices between non-obese PCOS and control groups. AUC, area under the curve; LAP, lipid accumulation product; PCOS, polycystic ovary syndrome; ROC, receiver operating characteristic; VAI, Visceral adiposity index. AUC values are shown in the legend; numerical AUCs and 95% CIs are provided in Table 5.

demonstrate a higher prevalence of atherogenic small, dense LDL particles [25].

To elucidate obesity-independent effects of PCOS on atherogenic burden, we stratified the PCOS cohort by BMI. As expected, IR indices differed significantly between obese and non-obese PCOS subgroups. HOMA-IR and METS-IR were higher, QUICKI and FG-IR were lower in the obese PCOS group, confirming obesity's contribution to worsening IR. Additionally, our comprehensive assessment of multiple IR indices—HOMA-IR, QUICKI, FG-IR, and METS-IR—consistently demonstrated significantly greater IR in non-obese PCOS compared to BMI-matched controls. This confirms that IR in PCOS extends beyond obesity-associated mechanisms, reflecting intrinsic metabolic dysfunction. Remarkably, pure atherogenic lipid indices showed no significant differences between obese and non-obese PCOS subgroups. We also observed that both pure atherogenic indices and IR-based cardiovascular risk indicators were significantly higher in non-obese PCOS compared to non-obese controls. Importantly, this adverse profile was not restricted to composite atherogenic indices; routine lipid measures, including TG, TC and non-HDL-C, were also significantly higher in non-obese women with PCOS than in non-obese controls, indicating that an atherogenic shift is detectable even with standard lipid panels. Collectively, these findings are compatible with the concept that the metabolic derangements driving atherogenic dyslipidemia in our cohort may stem predominantly from core PCOS pathophysiology—such as chronic hyperandrogenism, ovarian dysfunction and IR—rather than merely reflecting the consequences of excess adiposity. However, the cross-sectional design precludes definitive causal inferences.

Our ROC analysis identified LAP, VAI, and TyG as superior discriminators of atherogenic lipid indices in non-obese PCOS, with LAP demonstrating optimal overall performance (AUC=0.747). These composite indices integrate multiple metabolic components, potentially capturing cardiovascular risk more comprehensively than isolated lipid parameters.

LAP index is a simple marker of central lipid accumulation, and is strongly linked to increased cardiovascular risk—including in women with PCOS, both obese and non-obese [27-32]. Several large cohort studies indicate that higher LAP values are independently associated with an increased risk of cardiovascular events and mortality, surpassing BMI in certain populations [27-29]. Additionally, some studies have found the LAP index in PCOS to be strongly associated with impaired glucose tolerance, IR, and metabolic syndrome, which are key drivers of cardiovascular risk [30-32]. In our study, LAP index was found to be higher in the non-obese PCOS group than in the BMI-matched control group. In this context, the LAP value of 18.05 identified, within our cohort, non-obese women with PCOS who displayed a more adverse atherogenic profile using only routine anthropometric and lipid measurements.

VAI is a validated, sex-specific index to estimate visceral fat function and predict cardiometabolic risk, including CVD. Numerous studies show that higher VAI is independently associated with increased risk of CVD, heart failure, and metabolic syndrome in diverse populations [33-35]. Amato *et al*. proposed that the VAI could serve as a significant metric in both clinical settings and population-based research for evaluating cardiovascular risk in individuals with PCOS [36]. In our investigation, we observed that the VAI was significantly elevated in the non-obese PCOS cohort relative to the non-obese healthy control group. The analysis revealed a favorable discrimination performance, indicated by an AUC of 0.707, along with an optimal cut-off value of 1.15 in our non-obese PCOS population. These findings may imply a dysfunction of visceral adipose tissue, even among individuals maintaining a normal BMI.

The TyG index serves as an estimate of insulin resistance. A higher TyG index has been consistently associated with elevated CVD risk across diverse populations [9, 37]. In a prospective cohort, López-Jaramillo *et al*. reported that the TyG index was significantly associated with subsequent cardiovascular mortality, myocardial infarction, stroke, and type 2 DM [9]. In a meta-analysis of twelve cohort studies (six prospective and six retrospective) comprising 6,354,990 participants, Liu *et al*. identified a likely linear relationship between the TyG index and incident coronary artery disease and composite CVD [37]. However, direct evidence linking the TyG index to cardiovascular risk specifically in PCOS—in either obese or non-obese patients—is lacking. In our investigation, we found that the TyG index was significantly elevated in the non-obese PCOS cohort

compared to control participants. Moreover, the threshold of 8.29 identified in our ROC analysis demonstrated fair discriminatory ability (AUC = 0.701) with relatively high specificity (79%) in this sample.

It should be noted that AIP is essentially a logarithmic transformation of the TG/HDL-C ratio; thus, both indices preserve the same ranking of individuals and are expected to yield identical AUC values, as observed in our cohort.

A noteworthy point in our study is that, although LAP had the numerically highest AUC, the DeLong test did not demonstrate a statistically significant superiority of LAP over other indices, including VAI, TyG, AIP, TG/HDL-C ratio, LCI, Castelli indices, AC, and METS-IR. These findings suggest that multiple indices capturing overlapping aspects of IR and atherogenic dyslipidemia perform similarly and that index selection may be guided by clinical applicability rather than clear differences in AUC.

In the present single-center, cross-sectional study, the cut-off values derived from ROC analysis for LAP, VAI and TyG should be interpreted as exploratory and cohort-specific. Rather than representing universally applicable clinical thresholds, these values may indicate approximate levels at which non-obese women with PCOS in our population begin to exhibit a more adverse atherogenic profile compared with BMI-matched controls. Importantly, the performance of these indices and their corresponding thresholds is likely to vary across ethnic groups. Consequently, these indices and cut-points should not yet be used as stand-alone decision tools in clinical practice; instead, they may serve as supportive markers and starting points for future validation studies and risk-prediction models.

Strengths and Limitations

Several limitations warrant acknowledgment. The cross-sectional design precludes causal inference and assessment of temporal relationships between PCOS, atherogenic lipid indices, and cardiovascular outcomes. Longitudinal studies evaluating whether elevated indices predict incident cardiovascular events in non-obese PCOS would strengthen clinical applicability. Additionally, the study population consisted solely of women from one geographic location and ethnic background. This limitation reduces generalizability to other ethnicities, considering known racial and ethnic differences in PCOS phenotype and cardiovascular risk profiles. Validation of our identified thresholds in independent PCOS cohorts from diverse geographic and ethnic backgrounds would establish generalizability or enable the determination of ethnicity-specific cutoff values optimized for different populations.

CONCLUSION

This study demonstrates that non-obese women with PCOS exhibit significantly elevated atherogenic lipid and IR indices compared to BMI-matched healthy controls. Moreover, atherogenic dyslipidemia markers were similar in obese and non-obese groups in patients with PCOS. These findings are consistent with the notion that PCOS may independently contribute to cardiovascular risk through mechanisms extending beyond obesity-related metabolic dysfunction. Furthermore, the identified optimal thresholds for LAP (>18.05), VAI (>1.15), and TyG (>8.29) should be considered hypothesis-generating and specific to our cohort. Future research should validate these thresholds in diverse populations, assess their prognostic value for predicting cardiovascular events, and investigate whether targeted therapeutic interventions that improve these atherogenic indices lead to significant reductions in cardiometabolic risk among non-obese individuals with PCOS.

Ethics Approval and Consent to Participate

This study was approved by the Ordu University Faculty of Medicine Ethics Committee (Decision No: BAEK 2025/341; date: 17.10.2025). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Because the study was retrospective and no additional intervention was performed on the participants, the informed consent form was waived.

Data Availability

All data generated or analyzed during this study are included in this published article. The data that

support the findings of this study are available on request from the corresponding author, upon reasonable request.

Authors' Contribution

Study Conception: BA, MNB, GNE; Study Design: BA, MNB; Supervision: BA; Funding: N/A; Materials: BA, GNE; Data Collection and/or Processing: BA, GNE; Statistical Analysis and/or Data Interpretation: BA, MNB; Literature Review: BA, MNB, GNE; Manuscript Preparation: BA, MNB, GNE; and Critical Review: BA, MNB, GNE.

Conflict of Interest

The author(s) disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The author(s) disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgments

The findings of our research were presented as an abstract at the 2025 TEMD ENDOKURS Congress.

Generative Artificial Intelligence Statement

The author(s) declare that no artificial intelligence-based tools or applications were used during the preparation process of this manuscript. However, The "Grammarly" application was used to review the article in terms of English language corrections and grammar. The all content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

Editor's Note

All statements made in this article are solely those of the authors and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Helvaci N, Yildiz BO. Polycystic ovary syndrome as a metabolic disease. Nat Rev Endocrinol. 2025;21(4):230-244. doi: 10.1038/s41574-024-01057-w.
2. 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):2841-2855. doi: 10.1093/humrep/dew218.
3. Barbieri RL, Ehrmann DA. Clinical manifestations of polycystic ovary syndrome in adults. In: Harris MB, editor. UpToDate. Waltham, MA: UpToDate; Updated Aug 27, 2025. Accessed Oct 25, 2025. Available from: https://www.uptodate.com/contents/clinical-manifestations-of-polycystic-ovary-syndrome-in-adults
4. de Groot PC, Dekkers OM, Romijn JA, Dieben SW, Helmerhorst FM. PCOS, coronary heart disease, stroke and the influence of obesity: a systematic review and meta-analysis. Hum Reprod Update. 2011;17(4):495-500. doi: 10.1093/humupd/dmr001.
5. Wild RA. Polycystic ovary syndrome: a risk for coronary artery disease? Am J Obstet Gynecol. 2002;186(1):35-43. doi: 10.1067/mob.2002.119180.
6. Elci E, Kaya C, Cim N, Yildizhan R, Elci GG. Evaluation of cardiac risk marker levels in obese and non-obese patients with polycystic ovaries. Gynecol Endocrinol. 2017;33(1):43-47. doi: 10.1080/09513590.2016.1203893.
7. Choi YS, Yang HI, Cho S, et al. Serum asymmetric dimethylarginine, apelin, and tumor necrosis factor-α levels in non-obese women with polycystic ovary syndrome. Steroids. 2012;77(13):1352-1358. doi: 10.1016/j.steroids.2012.08.005.
8. Jabbour R, Ott J, Eppel W, Frigo P. Carotid intima-media thickness in polycystic ovary syndrome and its association with hormone and lipid profiles. PLoS One. 2020;15(4):e0232299. doi: 10.1371/journal.pone.0232299.
9. Lopez-Jaramillo P, Gomez-Arbelaez D, Martinez-Bello D, et al. Association of the triglyceride glucose index as a measure of insulin resistance with mortality and cardiovascular disease in populations from five continents (PURE study): a prospective cohort study. Lancet Healthy Longev. 2023;4(1):e23-e33. doi: 10.1016/S2666-7568(22)00247-1.
10. Darroudi S, Soflaee SS, Hosseini ZS, et al. The visceral adiposity index and lipid accumulation product as predictors of cardiovascular events in normal weight subjects. Clin Nutr ESPEN. 2022;52:190-197. doi: 10.1016/j.clnesp.2022.10.015.
11. Duan M, Zhao X, Li S, etbal. Metabolic score for insulin resistance (METS-IR) predicts all-cause and cardiovascular mortality in the general population: evidence from NHANES 2001-2018. Cardiovasc Diabetol. 2024;23(1):243. doi: 10.1186/s12933-024-02334-8.
12. Uysal E, Tammo O, Soylemez E, Incebıyık M, Filiz D, Alci M. Significance of measuring anthropometric and atherogenic indices in patients with polycystic ovary syndrome. BMC Endocr Disord. 2024;24(1):160. doi: 10.1186/s12902-024-01701-6.
13. van der Ham K, Louwers YV, Laven JSE. Cardiometabolic biomarkers in women with polycystic ovary syndrome. Fertil Steril. 2022;117(5):887-896. doi: 10.1016/j.fertnstert.2022.03.008.
14. Brończyk-Puzoń A, Jagielski P, Kulik-Kupka K, Koszowska A, Nowak J, Zubelewicz-Szkodzińska B. Usefulness of a new anthropometric indicator - VAI (Visceral Adiposity Index) in the evaluation of metabolic and hormonal disorders in women with polycystic ovary syndrome. Adv Clin Exp Med. 2017;26(5):825-828. doi: 10.17219/acem/61100.

15. Kałużna M, Czlapka-Matyasik M, Kompf P, et al. Lipid ratios and obesity indices are effective predictors of metabolic syndrome in women with polycystic ovary syndrome. Ther Adv Endocrinol Metab. 2022;13:20420188211066699. doi: 10.1177/20420188211066699.
16. 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. doi: 10.1016/j.fertnstert.2003.10.004.
17. Matthews DR, Hosker JP, Rudenski AS, Naylor BA, Treacher DF, Turner RC. Homeostasis model assessment: insulin resistance and beta-cell function from fasting plasma glucose and insulin concentrations in man. Diabetologia. 1985;28(7):412-419. doi: 10.1007/BF00280883.
18. Katz A, Nambi SS, Mather K, et al. Quantitative insulin sensitivity check index: a simple, accurate method for assessing insulin sensitivity in humans. J Clin Endocrinol Metab. 2000;85(7):2402-2410. doi: 10.1210/jcem.85.7.6661.
19. Castelli WP, Abbott RD, McNamara PM. Summary estimates of cholesterol used to predict coronary heart disease. Circulation. 1983;67(4):730-734. doi: 10.1161/01.cir.67.4.730.
20. Diamanti-Kandarakis E, Dunaif A. Insulin resistance and the polycystic ovary syndrome revisited: an update on mechanisms and implications. Endocr Rev. 2012;33(6):981-1030. doi: 10.1210/er.2011-1034.
21. Paradisi G, Steinberg HO, Hempfling A, et al. Polycystic ovary syndrome is associated with endothelial dysfunction. Circulation. 2001;103(10):1410-1415. doi: 10.1161/01.cir.103.10.1410.
22. Gomez JMD, VanHise K, Stachenfeld N, Chan JL, Merz NB, Shufelt C. Subclinical cardiovascular disease and polycystic ovary syndrome. Fertil Steril. 2022;117(5):912-923. doi: 10.1016/j.fertnstert.2022.02.028.
23. Usselman CW, Yarovinsky TO, Steele FE, et al. Androgens drive microvascular endothelial dysfunction in women with polycystic ovary syndrome: role of the endothelin B receptor. J Physiol. 2019;597(11):2853-2865. doi: 10.1113/JP277756.
24. Anagnostis P, Paparodis RD, Bosdou JK, et al. Risk of type 2 diabetes mellitus in polycystic ovary syndrome is associated with obesity: a meta-analysis of observational studies. Endocrine. 2021;74(2):245-253. doi: 10.1007/s12020-021-02801-2.
25. Phelan N, O'Connor A, Kyaw-Tun T, et al. Lipoprotein subclass patterns in women with polycystic ovary syndrome (PCOS) compared with equally insulin-resistant women without PCOS. J Clin Endocrinol Metab. 2010;95(8):3933-3939. doi: 10.1210/jc.2009-2444.
26. Rudnicka E, Suchta K, Grymowicz M, et al. Chronic Low Grade Inflammation in Pathogenesis of PCOS. Int J Mol Sci. 2021;22(7):3789. doi: 10.3390/ijms22073789.
27. Kyrou I, Panagiotakos DB, Kouli GM, et al. Lipid accumulation product in relation to 10-year cardiovascular disease incidence in Caucasian adults: The ATTICA study. Atherosclerosis. 2018;279:10-16. doi: 10.1016/j.atherosclerosis.2018.10.015.
28. Hosseinpanah F, Barzin M, Mirbolouk M, Abtahi H, Cheraghi L, Azizi F. Lipid accumulation product and incident cardiovascular events in a normal weight population: Tehran Lipid and Glucose Study. Eur J Prev Cardiol. 2016;23(2):187-193. doi: 10.1177/2047487314558771.
29. Ioachimescu AG, Brennan DM, Hoar BM, Hoogwerf BJ. The lipid accumulation product and all-cause mortality in patients at high cardiovascular risk: a PreCIS database study. Obesity (Silver Spring). 2010;18(9):1836-1844. doi: 10.1038/oby.2009.453.
30. Wehr E, Gruber HJ, Giuliani A, Möller R, Pieber TR, Obermayer-Pietsch B. The lipid accumulation product is associated with impaired glucose tolerance in PCOS women. J Clin Endocrinol Metab. 2011;96(6):E986-990. doi: 10.1210/jc.2011-0031.
31. Wiltgen D, Spritzer PM. Variation in metabolic and cardiovascular risk in women with different polycystic ovary syndrome phenotypes. Fertil Steril. 2010;94(6):2493-2496. doi: 10.1016/j.fertnstert.2010.02.015.
32. Macut D, Tziomalos K, Božić-Antić I, et al. Non-alcoholic fatty liver disease is associated with insulin resistance and lipid accumulation product in women with polycystic ovary syndrome. Hum Reprod. 2016;31(6):1347-1353. doi: 10.1093/humrep/dew076.
33. Yuan Y, Hu X, Jin J, et al. Transition of visceral adiposity index and risk of cardiovascular disease in middle-aged and older Chinese adults. Arch Gerontol Geriatr. 2024;121:105356. doi: 10.1016/j.archger.2024.105356.
34. Lazzer S, D'Alleva M, Isola M, et al. Cardiometabolic Index (CMI) and Visceral Adiposity Index (VAI) Highlight a Higher Risk of Metabolic Syndrome in Women with Severe Obesity. J Clin Med. 2023;12(9):3055. doi: 10.3390/jcm12093055.
35. Xu C, Guo Y, Zhang S, et al. Visceral adiposity index and the risk of heart failure, late-life cardiac structure, and function in ARIC study. Eur J Prev Cardiol. 2023;30(12):1182-1192. doi: 10.1093/eurjpc/zwad099.
36. Amato MC, Verghi M, Galluzzo A, Giordano C. The oligomenorrhoic phenotypes of polycystic ovary syndrome are characterized by a high visceral adiposity index: a likely condition of cardiometabolic risk. Hum Reprod. 2011;26(6):1486-1494. doi: 10.1093/humrep/der088.
37. Liu X, Tan Z, Huang Y, et al. Relationship between the triglyceride-glucose index and risk of cardiovascular diseases and mortality in the general population: a systematic review and meta-analysis. Cardiovasc Diabetol. 2022;21(1):124. doi: 10.1186/s12933-022-01546-0.

The European Research Journal 2026;12(7):752-759 Research Article

DOI: https://doi.org/10.18621/eurj.1840748 *Public Health*

Smartphone Addiction, Sleep Quality, and Temporomandibular Disorders Among Dental Students: A Cross-Sectional Analysis

Ezgi Sıla Taşkaldıran[1], Gülce Nil Varlıhan[1], Sevda Naghizadeh Asgari[1]

[1]*Department of Periodontology, Istanbul Aydın University, Faculty of Dentistry, Istanbul, Türkiye*

Abstract:

Objective: The rapid digitalization of daily life has made smartphone dependence a growing public health concern, particularly among university students. This study explored how smartphone addiction relates to sleep quality, temporomandibular disorder (TMD), and bruxism in dental students at different stages of training.

Methods: A total of 259 participants completed validated questionnaires assessing the Smartphone Addiction Scale–Short Version (SAS-SV), Pittsburgh Sleep Quality Index, bruxism, and TMD symptoms.

Results: Nearly half of the students exhibited signs of smartphone addiction. Higher addiction scores were significantly correlated with poorer sleep quality and greater TMD severity, particularly among preclinical and doctoral students, while no significant relationship was found between bruxism and addiction scores.

Conclusion: These results imply that excessive smartphone use may silently impact mental and orofacial health by causing stress, sleep disturbances, and postural strain. Promoting healthier technology habits and addressing digital dependence could be key to protecting the well-being of future dental professionals.

Keywords: Smartphone Addiction Scale, Pittsburgh Sleep Quality Index, Temporomandibular Disorders, Bruxism, Dental Students

The 21st century has witnessed a profound transformation in human behavior driven by the rapid evolution of digital technology. Smartphones, used by more than 6.6 billion people worldwide, have become essential tools for communication, learning, and organization [1, 2]. Despite their benefits, concerns have grown regarding excessive use and its impact on public health [3].

Smartphone addiction is often described as a behavioral form of dependence, marked by compulsive use, withdrawal symptoms, and difficulties in self-regulation [4]. Similar to non-substance addictions, it involves tolerance, preoccupation, and adverse effects on mental and physical health [4-6]. University students, particularly those on health-related degrees, seem to be particularly vulnerable. High usage rates are consistently associated with poor sleep, emotional distress and reduced academic performance [7].

Sleep problems are among the most consistent findings in studies on problematic smartphone use. Light exposure from screens suppresses melatonin and

Submitted: December 12, 2025 **Accepted:** January 17, 2026 **Published Online:** January 29, 2026

How to cite this article: Taşkaldıran ES, Varlıhan GN, Naghizadeh Asgari S. Smartphone Addiction, Sleep Quality, and Temporomandibular Disorders Among Dental Students: A Cross-Sectional Analysis. *Eur Res J.* 2026;12(7):752-759. doi: 10.18621/eurj.1840748

Corresponding author: Ezgi Sila Taskaldiran, PhD., Asisst. Prof., Phone: +90 212 411 30 00, E-mail: ezgisilataskaldiran@aydin.edu.tr

Available Online at https://www.eurj.org.tr

delays sleep onset [8, 9]. In addition, habits such as "revenge bedtime scrolling" cause individuals to delay sleep to gain personal time [10]. Over time, these patterns contribute to reduced cognitive performance, weakened immunity, and increased physiological stress [11]. Some evidence also links digital overuse with oral health problems, such as fatigue-related dysfunctions and changes in periodontal health [12].

Another issue concerns posture and musculoskeletal strain. The forward head posture commonly adopted during smartphone use increases the mechanical load on cervical and masticatory muscles, which may predispose users to temporomandibular disorders (TMDs) [13]. These disorders often coexist with stress and sleep disturbance. Bruxism, characterized by clenching or grinding of teeth, is also linked to emotional tension and irregular routines [14-16]. Health science students are particularly susceptible, as heavy academic demands and continuous smartphone use may impair their sleep, mood, and academic efficiency. Recent research increasingly emphasizes the reciprocal nature of these interactions. Behavioral and environmental factors such as digital overuse, poor posture, and insufficient rest appear to influence both psychological balance and musculoskeletal stability [17]. Students in health-related fields, particularly dental and medical programs, occupy a unique position within this landscape. Their long study hours, clinical responsibilities, and dependence on digital tools can lead to fatigue, postural discomfort, and irregular sleep patterns [3, 7, 11].

In this context, the current study was designed based on the hypothesis that there may be differences in the levels of smartphone addiction among dental students at different stages of their academic journey. The study aimed to determine the degree of smartphone addiction in this population and to examine its associations with sleep quality, TMD, and bruxism, and thus the impact of digital behaviors in this population.

METHODS

Study Design and Participants

This cross-sectional study was conducted among students enrolled at the Istanbul Aydın University Faculty of Dentistry during the 2024–2025 academic year. Ethical approval for the study was granted by the local ethics committee (Non-Interventional Clinical Research Ethics Committee (2024/66). All study procedures were performed in accordance with the ethical principles of the Declaration of Helsinki.

In order to ensure adequate statistical validity, the preliminary study power analysis indicated that the required sample size was obtained with a significance level of $\alpha = 0.05$ and a statistical power of $1\text{-}\beta = 0.80$. In consideration of the potential data loss, a total of 270 individuals were included in the study.

The inclusion criteria comprised undergraduate and postgraduate dental students enrolled in the Istanbul Aydin University, Faculty of Dentistry, who were literate in Turkish and voluntarily provided written informed consent. All participants who met the specified criteria and had completed the questionnaire were initially included in the study. Following the collection of data, responses containing missing or inconsistent information were excluded from the final analysis to ensure data quality.

The study population, consisting of 270 participants, was divided into three academic groups: the preclinical group (Group P; 3rd-year students), the clinical group (Group C; 4th- and 5th-year students), and the doctoral group (Group D; PhD students), with participants distributed equally across groups. In order to maintain standardization in academic workload and stress-related conditions, all data were collected during the final examination period at the end of the year.

Data Collection

The data were collected using a structured, self-administered questionnaire, which was distributed in classroom settings under the supervision of a researcher from the study team (SNA). Before participation, all subjects were informed about the study's purpose and confidentiality principles, and provided written informed consent. The survey was designed to require approximately 15–20 minutes to complete, and participants were permitted to review their responses prior to submission. The questionnaire consisted of four main sections: a demographic section collecting age, gender, and academic level, followed by four validated questionnaires evaluating smartphone addiction, sleep quality, bruxism, and TMDs.

The assessment of smartphone addiction was conducted utilizing the Smartphone Addiction Scale–Short Version (SAS-SV), which was originally developed by Kwon *et al*. [4] and subsequently validated for reliability and construct validity in Turkish [18]. The scale under consideration consists of 33 items, which are to be rated on a six-point Likert scale ranging from 1 (strongly disagree) to 6 (strongly agree). Scores on this measure are positively correlated with the severity of symptoms, indicating a smartphone addiction.

The evaluation of sleep quality was conducted utilizing the Pittsburgh Sleep Quality Index (PSQI) [19, 20]. The questionnaire under consideration comprises 19 self-rated items, which are grouped into seven components as follows: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction. Each component is assigned a score ranging from 0 to 3, resulting in a global score that extends from 0 to 21. Scores of 5 and above are indicative of poor sleep quality.

Bruxism was assessed using the Turkish version [21] of the Bruxism Assessment Index (BAI), which was originally developed by Lobbezoo *et al*. [14]. The questionnaire evaluates both awake and sleep bruxism through self-reported items answered as "Yes", "Sometimes", or "No". Higher total scores indicate a greater frequency and severity of bruxism-related behaviors.

The final part of the questionnaire assessed TMD using the validated Turkish version of the Fonseca Anamnestic Index (FAI) [22, 23]. This includes 10 self-reported items, which are scored as 0 (no), 5 (sometimes), or 10 (yes). This results in a total score ranging from 0 to 100. TMD severity is then classified as mild (20–40), moderate (45–65), or severe (70–100) based on these scores.

After data collection, 11 questionnaires containing missing or inconsistent responses were excluded in accordance with the predefined exclusion criteria. The final sample, therefore, consisted of 259 participants: 90 from Group P, 87 from Group C, and 82 from Group D.

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics for Windows, Version 29.0 (IBM Corp., Armonk, NY, USA). The Kolmogorov–Smirnov test was used to assess the normality of data distribution, indicating that the variables did not follow a normal distribution. Therefore, nonparametric tests were applied. Descriptive statistics were calculated for all variables, including minimum, maximum, mean, standard deviation, median, and frequency. Quantitative variables were compared among the three academic groups using the Kruskal–Wallis H test, followed by Dunn's post hoc test for pairwise comparisons. The Mann–Whitney U test was used for comparisons between two independent groups. Qualitative variables were analysed using the Chi-square test, Fisher–Freeman–Halton Exact test, or Continuity (Yates) correction as appropriate. Correlations between continuous or ordinal variables were examined using Spearman's rank-order (rho) correlation analysis. All statistical tests were two-tailed, and $P<0.05$ was considered statistically significant.

RESULTS

Demographic Characteristics

The 259 participants were aged between 20 and 47 years (24.92±3.59), with 57.5% (n=149) being female and 42.5% (n=110) being male.

Of the participants in Group C, 71.3% (n=62) were 4th-year students and 28.7% (n=25) were 5th-year students. In Group D, 22.0% (n=18) were in their 1st year, 50.0% (n=41) in the 2nd year, 24.4% (n=20) in the 3rd year, and 3.7% (n=3) in the 4th year of the doctoral program.

A statistically significant difference in mean age was observed among the groups, with Group D exhibiting a higher mean age than both Group P and Group C ($P= 0.001$ for both). Group C also showed a higher mean age than Group P ($P= 0.001$). Gender distribution did not differ significantly across the groups ($P > 0.05$) (Table 1).

Survey Data

No statistically significant difference was found in SAS-SV scores among the groups ($P = 0.181$). Similarly, no significant difference was observed between 4th- and 5th-year students within Group C

TABLE 1. Comparison of Age and Gender Distribution Among Groups

		Group P (n=90)	Group C (n=87)	Group D (n=82)	P-value
Age (years)		22.39±2.97 (22)	24.30±2.66 (24)	28.36±3.04 (28)	**0.001[a]**
Gender	Female	50 (55.6%)	50 (57.5%)	49 (59.8%)	0.856[b]
	Male	40 (44.4%)	37 (42.5%)	33 (40.2%)	

Data are shown as mean±standard deviation (median) or n (%) where appropriate. Group P, Preclinical Group; Group C, Clinical Group; Group D, Doctoral Group.

[a]Kruskal Wallis test, [b]Ki-kare test. Statistically significant P-value is shown in bold.

(P=0.331).

PSQI scores differed significantly among the groups (P=0.002). Post hoc analysis revealed that Group C had significantly higher PSQI scores compared to Group D (P=0.001), while no significant differences were found between the other groups (P>0.05). There was also no significant difference in PSQI scores between 4th- and 5th-year students within Group C (P>0.05).

TMD scores did not show a statistically significant difference among the groups (P=0.102). However, 4th-year students demonstrated significantly higher TMD scores compared to 5th-year students within Group C (P=0.008).

T he prevalence of smartphone addiction and bruxism did not differ significantly among the groups (P=0.148). Likewise, no significant differences were observed between 4th- and 5th-year students in either smartphone addiction or bruxism prevalence (P=0.129).

No statistically significant differences were found among academic years within Group D in SAS-SV,

TABLE 2. Comparison of Survey Parameters Among Groups

		Group P (n=90)	Group C (n=87)	Group D (n=82)	P-value
SAS-SV score		33.33±9.68 32 (13-55)	31.75±9.94 30 (11-60)	30.84±10.37 30 (10-60)	0.181[a]
PSQI score		8.02±3.54 8 (1-17)	8.75±2.93 9 (3-14)	7.11±3.14 7 (2-17)	**0.002[a]**
TMD score		41.22±24.59 40 (0-100)	38.85±24.93 35 (0-100)	33.29±24.5 30 (0-100)	0.102[a]
Smartphone addiction prevalence	No	45 (50%)	51 (58.6%)	53 (64.6%)	0.148[b]
	Yes	45 (50%)	36 (41.4%)	29 (35.4%)	
Bruxism prevalence	No	32 (35.6%)	27 (31%)	36 (43.9%)	0.214[b]
	Yes	58 (64.4%)	60 (69%)	46 (56.1%)	
TMD severity	None	17 (18.9%)	22 (25.3%)	24 (29.3%)	0.443[b]
	Mild	35 (38.9%)	31 (35.6%)	36 (43.9%)	
	Moderate	23 (25.6%)	21 (24.1%)	14 (17.1%)	
	Severe	15 (16.7%)	13 (14.9%)	8 (9.8%)	

Data are shown as mean±standard deviation or median (minimum-maximum) or n (%) where appropriate. SAS-SV, Smartphone addiction scale-short version; PSQI, Pittsburgh sleep quality index; TMD, temporomandibular disorder. Group P, Preclinical Group; Group C, Clinical Group; Group D, Doctoral Group.

[a]Kruskal Wallis test, [b]Ki-kare test. Statistically significant P-value is shown in bold.

PSQI, or TMD scores, nor in the prevalence of smartphone addiction, bruxism, or TMD severity (P=0.653, P=0.068, P=0.539, P=0.201, P=0.499, and P=0.443, respectively). All survey data are presented in Table 2.

Correlation Analysis

In the overall analysis, SAS-SV scores demonstrated statistically significant positive correlations with PSQI scores (r = 0.152, P=0.014) and with TMD scores (r = 0.186, P=0.003). However, the statistical analysis did not reveal a significant association between SAS-SV scores and bruxism (P>0.05). When analysed by group, a statistically significant positive correlation was found between SAS-SV and PSQI scores in Group D (r = 0.272, P=0.014) and between SAS-SV and TMD scores in Group P (r = 0.272, P=0.010). No statistically significant correlations were detected within Group C.

DISCUSSION

This study examined the level of smartphone addiction among dental students at different academic stages, as well as its relationship with sleep quality, bruxism, and TMD. A considerable percentage of students in each academic group - ranging from one-third to nearly one-half - met the criteria for smartphone addiction, reflecting a notable prevalence of problematic smartphone use in this population. The findings also revealed statistically significant positive correlations between smartphone addiction, poor sleep quality, and TMD scores, suggesting that excessive smartphone use may be associated with both psychological and somatic consequences. These findings are consistent with an expanding body of evidence underscoring the biopsychosocial consequences of problematic smartphone use [3, 24-26].

The observed association between smartphone addiction and poor sleep quality in the present study is consistent with previous research. Excessive smartphone use, particularly during nighttime hours, has been demonstrated to delay sleep onset, shorten sleep duration, and impair overall sleep efficiency due to blue light exposure and behavioral hyperarousal [1, 19, 27]. Nikolic *et al*. [7] reported that 21.7% of medical students met the criteria for smartphone addiction, which was significantly associated with poorer sleep quality, anxiety, and depression. In line with these findings, Demirci *et al*. [27] and Joshi [28] also observed that problematic smartphone use negatively affects both sleep hygiene and psychological well-being among university students. The present findings among dental students support these observations, highlighting that this population-characterized by high academic workload and stress, may be particularly vulnerable to sleep disturbances associated with excessive smartphone use. Consistently, the significant positive correlation between SAS-SV and PSQI scores observed in our study further reinforces the close link between digital dependence and poor sleep quality. This association was particularly evident among doctoral students, which may reflect the impact of increased academic workload, prolonged screen exposure related to research activities, and heightened psychological stress commonly experienced during advanced training. This relationship can also be explained by the behavioral concept of digital dependence. Recent studies have described "revenge bedtime scrolling" as a form of self-regulation in which individuals intentionally delay sleep to reclaim personal time lost to academic or work responsibilities. Such patterns are likely to be prevalent among dental students, who often have long study hours, limited leisure time, and a high cognitive load. Consequently, continuous nighttime engagement with digital devices may perpetuate a cycle of delayed sleep, reduced sleep quality, and increased daytime fatigue.

TMDs and bruxism are multifactorial conditions influenced by physiological, psychological, and behavioral factors. In recent years, increasing attention has been directed toward the potential contribution of digital behavior to musculoskeletal dysfunctions. de Jesus Correia *et al*. [13] reported that smartphone addiction is associated with musculoskeletal pain in the neck, shoulders, and upper limbs, primarily due to prolonged flexed posture and static muscle activation. Maintaining a forward head position for extended periods can increase strain on the cervical and masticatory muscles, thereby predisposing individuals to TMD symptoms [29]. In the present study, SAS-SV scores were significantly correlated with TMD scores, particularly among preclinical students. This

association may be attributed to posture-related strain compounded by academic stress. Kee *et al*. [16] similarly reported that sustained postural imbalance and repetitive screen-based activities alter temporomandibular biomechanics, exacerbating TMD-related discomfort. Moreover, psychological tension arising from excessive digital engagement may manifest as awake or sleep bruxism [14, 15]. Lobbezoo *et al*. [14] highlighted psychological stress as a principal etiological factor for bruxism, while Manfredini *et al*. [15] suggested that certain behavioral habits, including prolonged screen exposure, may contribute to increased parafunctional activity.

Smartphone addiction shares several neurobiological and behavioral features with other forms of behavioral addiction, including dopaminergic reinforcement and impaired impulse control [30]. Previous research has consistently reported strong associations between smartphone addiction, depression, anxiety, and stress [7, 25, 30]. Alageel *et al*. [31] further demonstrated that postgraduate students with higher levels of smartphone addiction were twice as likely to experience insomnia and exhibited significantly higher depression scores.

The weak but significant correlations observed in the present study between smartphone addiction, poor sleep quality, and TMD can be interpreted within this psychosomatic framework. Chronic stress and prolonged digital exposure may increase muscle tension, disrupt sleep regulation, and impair physiological recovery, thereby creating a self-perpetuating cycle of behavioral and somatic strain [29, 32].

The results of this study have important implications for health promotion among university students. Previous authors have emphasized smartphone addiction as an emerging public health concern due to its detrimental effects on mental health, sleep, and posture [3, 10, 32]. Loleska and Pop-Jordanova [3] argued that problematic smartphone use should be recognized as a behavioral addiction with measurable physiological consequences. Given that dental and medical students are among the heaviest smartphone users for both academic and social purposes, targeted preventive interventions are warranted. Educational programs should promote ergonomic awareness, digital hygiene, and stress management strategies. Additionally, structured breaks, posture training, and screen time regulation may help reduce the risk of TMD and sleep disturbances in this population [16, 19].

Strengths and Limitations

To the best of our knowledge, this study is the first to measure smartphone addiction among dental school students at different academic levels. The study also assessed sleep quality, TMD, and the presence of bruxism using valid scales. The presented study, which focuses on smartphone addiction as an important public health concern, has certain limitations in addition to these strengths. Its cross-sectional design does not allow for establishing causal relationships between smartphone addiction and the observed physiological or behavioral outcomes. All data were obtained through self-reported questionnaires, which may have introduced recall or social desirability bias. Furthermore, the relatively limited sample size may have affected the generalizability of the findings and reduced statistical power, particularly in subgroup comparisons among preclinical, clinical, and doctoral students. Future studies with larger and more diverse samples, employing longitudinal or interventional designs, are recommended to clarify causal mechanisms and to assess the effectiveness of preventive and behavioral strategies aimed at reducing smartphone-related health risks among students in health sciences.

CONCLUSION

This study highlights the significant associations between smartphone addiction, sleep quality, and TMD among dental students. Although group differences were not statistically significant, the observed correlations suggest that excessive smartphone use may indirectly influence orofacial health through pathways involving stress, poor sleep, and postural strain. Given the widespread use of smartphones in academic environments, increasing awareness of healthy digital habits, proper posture, and sleep hygiene is essential. Future studies employing longitudinal and experimental designs are needed to elucidate causal relationships and to develop targeted interventions aimed at minimizing the health risks associated with problematic smartphone use.

Ethics Approval and Consent to Participate

The study was approved by the Istanbul Aydın Unıversity Non-Interventional Clinical Research Ethics Committee (Decision no.: 2024/66, date: 31.07.2024). All procedures performed during data collection, review of patient records, and study implementation complied with the ethical standards of the institutional and/or national research committee and with the 1964 Declaration of Helsinki and its subsequent amendments. Written informed consent was obtained from all individual participants included in the study.

Data Availability

All data generated or analyzed during this study are included in this published article. The data that support the findings of this study are available on request from the corresponding author, upon reasonable request.

Authors' Contribution

Study Conception: EST, GNV; Study Design: EST, GNV; Supervision: EST, GNV; Data Collection and/or Processing: SNA; Statistical Analysis and/or Data Interpretation: EST, GNV; Literature Review: EST, GNV, SNA; Manuscript Preparation: EST, GNV, SNA; and Critical Review: EST, GNV.

Conflict of Interest

The author(s) disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The author(s) disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgment

The authors would like to thank the all participating patiens involved.

Generative Artificial Intelligence Statement

The author(s) declare that no artificial intelligence-based tools or applications were used during the preparation process of this manuscript. The all content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

Editor's Note

All statements made in this article are solely those of the authors and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Leow MQH, Chiang J, Chua TJX, Wang S, Tan NC. The relationship between smartphone addiction and sleep among medical students: A systematic review and meta-analysis. PLoS One. 2023;18(9):e0290724. doi: 10.1371/journal.pone.0290724.
2. Statista Research Department. Number of smartphone users worldwide from 2016 to 2028. Statista. April 2024. Available from: https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/
3. Loleska S, Pop-Jordanova N. Is Smartphone Addiction in the Younger Population a Public Health Problem? Pril (Makedon Akad Nauk Umet Odd Med Nauki). 2021;42(3):29-36. doi: 10.2478/prilozi-2021-0032.
4. Panova T, Carbonell X. Is smartphone addiction really an addiction? J Behav Addict. 2018;7(2):252-259. doi: 10.1556/2006.7.2018.49.
5. Zou Z, Wang H, d'Oleire Uquillas F, Wang X, Ding J, Chen H. Definition of substance and non-substance addiction. In: Zou Z, ed. Substance and Non-substance Addiction. Singapore: Springer; 2017: pp. 21-41. doi: 10.1007/978-981-10-5562-1_2.
6. Lin YH, Lin YC, Lee YH, Lin PH, Lin SH, Chang LR. Proposed diagnostic criteria for smartphone addiction. PLoS One. 2016;11(11):e0163010. doi: 10.1371/journal.pone.0163010.
7. Nikolic A, Bukurov B, Kocic I, *et al*. Smartphone addiction, sleep quality, depression, anxiety, and stress among medical students. Front Public Health. 2023;11:1252371. doi: 10.3389/fpubh.2023.1252371.
8. Chang AM, Aeschbach D, Duffy JF, Czeisler CA. Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. Proc Natl Acad Sci U S A. 2015;112(4):1232-1237. doi: 10.1073/pnas.1418490112.
9. Exelmans L, Van den Bulck J. Bedtime mobile phone use and sleep in adults. Soc Sci Med. 2016;148:93-101. doi: 10.1016/j.socscimed.2015.11.037.
10. Kroese FM, De Ridder DT, Evers C, Adriaanse MA. Bedtime procrastination: introducing a new area of procrastination. Front Psychol. 2014;5:611. doi: 10.3389/fpsyg.2014.00611.
11. Irwin MR. Sleep and inflammation: Partners in sickness and in health. Nat Rev Immunol. 2019;19(11):702-715. doi: 10.1038/s41577-019-0190-z.
12. Aksaka N, Şahinkaya S, Yay E. Influence of digital behavior and sleep quality on periodontal status in adolescents: A cross-sectional study. BMC Oral Health. 2025;25(1):1374. doi: 10.1186/s12903-025-06749-x.

13. de Jesus Correia F, Soares JB, dos Anjos Matos R, Pithon KR, Ferreira LN, de Assunção PL. Smartphone addiction, musculoskeletal pain and functionality in university students: An observational study. Psychol Health Med. 2024;29(2):286-296. doi: 10.1080/13548506.2023.2176893.
14. Lobbezoo F, Ahlberg J, Glaros AG, Kato T, Koyano K, Lavigne GJ. Bruxism defined and graded: An international consensus. J Oral Rehabil. 2013;40(1):2-4. doi: 10.1111/joor.12011.
15. Manfredini D, Winocur E, Guarda-Nardini L, Paesani D, Lobbezoo F. Epidemiology of bruxism in adults: A systematic review of the literature. J Orofac Pain. 2013;27(2):99-110. doi: 10.11607/jop.921.
16. Kee IK, Byun JS, Jung JK, Choi JK. The presence of altered craniocervical posture and mobility in smartphone-addicted teenagers with temporomandibular disorders. J Phys Ther Sci. 2016;28(2):339-346. doi: 10.1589/jpts.28.339.
17. Osorio-Molina C, Martos-Cabrera MB, Membrive-Jiménez MJ, Vargas-Roman K, Suleiman-Martos N, Ortega-Campos E. Smartphone addiction, risk factors and its adverse effects in nursing students: A systematic review and meta-analysis. Nurse Educ Today. 2021;98:104741. doi: 10.1016/j.nedt.2020.104741.
18. Noyan CO, Darçin AE, Nurmedov S, Yilmaz O, Dilbaz N. Akıllı Telefon Bağımlılığı Ölçeğinin Kısa Formunun üniversite öğrencilerindeTürkçe geçerlilik ve güvenilirlik çalışması 8Validity and reliability of the Turkish version of the Smartphone Addiction Scale-Short Version among university students]. Anadolu Psikiyatri Derg. 2015;16(Suppl 1):73-81. doi: 10.5455/apd.176101. [Article in Turkish]
19. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: A new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193-213. doi: 10.1016/0165-1781(89)90047-4.
20. Ağargün MY, Kara H, Anlar Ö. Pittsburgh uyku kalitesi indeksinin geçerliği ve güvenirliği [The Validity and Reliability of the Pittsburgh Sleep Quality Index]. Turk Psikiyatri Derg. 1996;7(2):107-115. [Article in Turkish]
21. Kaya M, Koroglu A, Sahin O. The relationship of psychological status and sociodemographic factors with bruxism among undergraduate dental students: A national survey. Niger J Clin Pract. 2022 Jun;25(6):944-950. doi: 10.4103/njcp.njcp_1980_21.
22. Fonseca DM, Bonfante G, Valle AL, Freitas SFT. Diagnóstico pela anamnese da disfunção craniomandibular. Rev Gaucha Odontol. 1994;42(1):23-28.
23. Kaynak BA, Taş S, Salkın Y. The accuracy and reliability of the Turkish version of the Fonseca anamnestic index in temporomandibular disorders. Cranio. 2023;41(1):78-83. doi: 10.1080/08869634.2020.1812808.
24. Griffiths M. A 'components' model of addiction within a biopsychosocial framework. J Subst Use. 2005;10(4):191-197. doi: 10.1080/14659890500114359.
25. De-Sola Gutiérrez J, Rodríguez de Fonseca F, Rubio G. Cell-phone addiction: A review. Front Psychiatry. 2016;7:175. doi: 10.3389/fpsyt.2016.00175.
26. Elhai JD, Dvorak RD, Levine JC, Hall BJ. Problematic smartphone use: A conceptual overview and systematic review of relations with anxiety and depression psychopathology. J Affect Disord. 2017;207:251-259. doi: 10.1016/j.jad.2016.08.030.
27. Demirci K, Akgönül M, Akpinar A. Relationship of smartphone use severity with sleep quality, depression, and anxiety in university students. J Behav Addict. 2015;4(2):85-92. doi: 10.1016/j.jad.2016.08.030.
28. Joshi SC. Sleep latency and sleep disturbances mediate the association between nighttime cell phone use and psychological well-being in college students. Sleep Biol Rhythms. 2022;20(3):431-443. doi: 10.1007/s41105-022-00388-3.
29. Regiani Bueno G, Garcia LF, Marques Gomes Bertolini SM, Rodrigues Lucena TF. The head down generation: Musculoskeletal symptoms and the use of smartphones among young university students. Telemed E Health. 2019;25(11):1049-1056. doi: 10.1089/tmj.2018.0231.
30. Thomée S, Härenstam A, Hagberg M. Mobile phone use and stress, sleep disturbances, and symptoms of depression among young adults: A prospective cohort study. BMC Public Health. 2011;11:66. doi: 10.1186/1471-2458-11-66.
31. Alageel AA, Alyahya RA, Bahatheq YA, et al. Smartphone addiction and associated factors among postgraduate students in an Arabic sample: A cross-sectional study. BMC Psychiatry. 2021;21(1):302. doi: 10.1186/s12888-021-03285-0.
32. Billieux J. Problematic use of the mobile phone: A literature review and a pathways model. Curr Psychiatry Rev. 2012;8(4):299-307. doi: 10.2174/157340012803520522.

The European Research Journal 2026;12(7):760-769 Research Article

DOI: https://doi.org/10.18621/eurj.1829545 *Public Health*

Awareness, Knowledge and Consumption of Postbiotics Among Students at Two Different Universities in Türkiye: A Cross-Sectional Survey

Ayşe Nur Kahve[1], Yağmur Yıldız[1]

[1]*Department of Field Sports and Health, Aksaray University, Aksaray, Türkiye*

Abstract:

Objective: Postbiotics, defined as non-viable microbial cells or their components that provide health benefits, are emerging as an important topic in functional food research. However, awareness and consumption among young adults remain limited. This study aimed to evaluate the awareness and knowledge levels of postbiotics among university students.

Methods: The study population included 346 students who completed a questionnaire consisting of 17 demographic and lifestyle questions and a 20-item postbiotic knowledge scale adapted from a validated tool. Descriptive statistics were used to analyze the data.

Results: It was revealed that 86.1% (n=298) of the students included in this study had never heard of postbiotics and 96.2% (n=333) had never used postbiotic supplements. Overall, 48.8% (n=169) of participants had a poor, 39.0% (n=135) a moderate, and 11.8% (n=41) a good knowledge level. Female students demonstrated significantly higher knowledge scores than males ($P<0.05$).

Conclusion: The findings indicate that university students have limited awareness and consumption of postbiotics. Contributing factors may include insufficient promotion, inadequate understanding of health benefits, and limited market availability of postbiotic products. To improve awareness, it is recommended that educational activities be organized in universities to enhance students' understanding of postbiotics and encourage their inclusion in health-promoting dietary habits.

Keywords: Functional Foods, Health Behavior, Knowledge, Postbiotics, Students

Functional foods can be defined as products specifically designed to meet any requirements in terms of functionality, nutrition, suitability, and medicinal properties [1]. The importance of biotics in the development of functional foods is widely recognized. Probiotics in supplement form are the most prominent component among functional foods in terms of their health effects. It has long been known that non-viable microorganisms, their cellular components, and metabolites can also influence health [2]. A variety of different terms, such as non-viable probiotics, heat-killed probiotics, cell lysates, paraprobiotics, and postbiotics, have begun to be addressed in various studies [3]. The term postbiotic originates from the Greek words "post," meaning after, and "bios," meaning life. It belongs to the

Submitted: November 24, 2025 **Accepted:** January 28, 2026 **Published Online:** January 31, 2026

How to cite this article: Kahve AN, Yıldız Y. Awareness, Knowledge and Consumption of Postbiotics Among Students at Two Different Universities in Türkiye: A Cross-Sectional Survey. *Eur Res J.* 2026;12(7):760-769. doi: 10.18621/eurj.1829545

Corresponding author: Ayşe Nur Kahve, PhD., Phone: +90 382 288 32 57, E-mail: aysenurkahve@aksaray.edu.tr

Available Online at https://www.eurj.org.tr

broader "biotic" family encompassing probiotics, prebiotics, synbiotics, and postbiotics all of which are associated with microorganisms or their substrates. Accordingly, postbiotics are defined as products that remain once microorganisms are no longer viable, meaning they are dead, inactive, or non-living [2]. These substances generally consist of a heterogeneous mixture of microbial cell components and metabolites, including teichoic acids, exopolysaccharides, peptidoglycans, bacteriocins, among others.

The effectiveness of postbiotics is primarily driven by three mechanisms: providing protection against pathogens, strengthening the epithelial barrier, and regulating both inflammatory and immune processes [4]. At present, their use extends beyond the field of fermented foods, as they are also being explored as a potential therapeutic option for various health issues, especially gastrointestinal problems like bloating and diarrhea. Consequently, postbiotics are expected to play a pivotal role in complementing probiotics and advancing the broader health industry [5]. Although the mechanisms related to the health-improving effects of postbiotics are not fully understood, they are believed to positively influence microbiota homeostasis and/or signaling pathways in the host. It has been reported that postbiotics can provide potential effects such as anti-inflammatory, antimicrobial, anti-obesity, immunomodulatory, anti-cancer, antihypertensive, antioxidant, and hypocholesterolemic effects [6, 7].

Recent studies have highlighted the potential health benefits of postbiotics across various physiological systems. Postbiotics have been associated with improved gut barrier integrity, modulation of immune responses, and regulation of inflammatory pathways [8]. In addition, evidence suggests that postbiotics may exert beneficial effects in conditions such as irritable bowel syndrome, metabolic disorders, obesity, and infections by enhancing host–microbiota interactions [9]. Compared to live probiotics, postbiotics offer advantages including improved safety, longer shelf life, and stability, making them promising candidates for functional food and therapeutic applications [10].

A review of the literature shows that awareness of the postbiotic concept is increasing, and there has been a rise in scientific studies conducted in recent years [11, 12]. However, there is a lack of sufficient studies indicating the consumption status of postbiotics and the knowledge levels within the community. This study aims to evaluate the postbiotic knowledge level and consumption status of postbiotics among university students.

METHODS

The research is a cross-sectional study designed to determine university students' postbiotic knowledge levels and consumption.

Population and Sample of the Study

The study population consisted of undergraduate students enrolled in the Faculties of Health Sciences, Tourism, and Sports Sciences at Aksaray University and Gazi University during the 2023-2024 academic year. A convenience sampling method was used due to voluntary participation and accessibility constraints. Students who met the inclusion criteria and agreed to participate were included in the study. Inclusion criteria were being aged 18 years or older and being enrolled in one of the specified faculties during the data collection period. Exclusion criteria included incomplete questionnaires and refusal to participate. The minimum required sample size was determined by reviewing similar cross-sectional studies conducted among university students investigating nutrition-related knowledge and behaviors. [13, 14]. Based on these studies, a sample size exceeding 300 participants was considered sufficient to ensure adequate representation and statistical power. A total of 346 students participated in the study. Participants were recruited proportionally from the Faculties of Health Sciences, Tourism, and Sports Sciences to reflect faculty-based distribution at both universities.

Data Collection Tools

A questionnaire consisting of 17 questions about gender, height, age, chronic diseases, family income, and the frequency of probiotic food consumption, along with a 20-item scale assessing postbiotic knowledge, was administered to university students. The postbiotic knowledge level was modeled from a validated scale by Batmaz [15]. The Cronbach's alpha

reliability coefficient of the modeled scale was found to be 0.72. The items in the modeled scale were developed by the team members to gather information about postbiotic knowledge and preferences, and they were tested for validity among several microbiologists and participants before the start of the study. The questionnaire was administered by evaluating team members to ensure that it was presented and interpreted in a similar manner. The questions in this section were evaluated using a 5-point Likert scale, with response options ranging from "strongly disagree," "disagree," "neutral," "agree," to "strongly agree." The maximum score that can be obtained under the knowledge level category is 80. The scores obtained from the scale were categorized as poor, moderate, good, and very good. A score of 65 or higher from the section was considered very good. The frequency of consumption of foods that could increase postbiotic production among students was measured using a modified food frequency questionnaire. When determining the frequency of consumption of foods that could increase postbiotic production, certain references were taken into account. Since postbiotics can be obtained from probiotics or their inactivation, and because postbiotics are produced biotechnologically through fermentation, yogurt, kefir, cabbage, and pickled vegetables have been identified as relevant foods.

Ethical Aspect of the Research

The research was conducted in accordance with ethical guidelines, and all procedures were carried out in compliance with the principles of the Helsinki Declaration. Ethical approval for the study was obtained from the Aksaray University Human Research Ethics Committee (number: 2024/01-41). Before starting the research, students were given the necessary explanations about the study and its procedures. Afterward, verbal consent was obtained from the students, and the questionnaire was administered.

Statistical Analysis

The statistical analysis of the data obtained in the study was conducted using SPSS 26 (IBM SPSS, IBM Corporation, USA) software. Descriptive statistics, including frequencies and percentages, were used to present the demographic data and the postbiotic knowledge levels of university students. Some percentage rates presented in the tables were calculated based on the total number of participants who responded to that question. In other sections of the study, Chi- squared test, t-tests and one-way analysis of variance (ANOVA) were conducted based on the characteristics of the variables. In the one-way analysis of variance (ANOVA), Tukey analysis was used for significance testing. In the tables, the arithmetic mean

TABLE 1. Basic Information About Students

	Numbers (n)	Total (%)
Gender		
Female	173	50.0
Male	173	50.0
Departments		
Recreation	18	5.2
Nursing	42	12.1
Recreation anagement	52	15.0
Physical education eaching	37	10.7
Coaching education	152	43.9
Sports management	45	13.1
Class		
1	55	15.9
2	133	38.4
3	57	16.5
4	101	29.2
Chronic Disease		
Yes	12	3.5
No	334	96.5
Family or personal financial status		
Poor	24	6.9
Moderate	209	60.4
Good	106	30.6
Very good	7	2.1
	Average	**Min-Max**
Age (years)	21.55	17-42
BMI (kg/m^2)	22.26	15-41

BMI, body mass index; Min, minimum; Max, maximum

(x), standard deviation (sd), and P-value are provided. A P-value less than 0.05 was considered statistically significant for testing differences.

RESULTS

Sociodemographic Characteristics of Students

The mean age of the 346 participants was 21.55 years (range: 17-42). The gender distribution was equal, with 173 (50.0%, n=173) females and 173 (50.0%, n=173) males. Regarding academic disciplines, 12.1% (n=42) were enrolled in Health Sciences, 72.8% (n=252) in Sports Sciences, and 15.0% (n=52) in Tourism-related programs. Table 1 summarizes the demographic characteristics.

Students Postbiotic Consumption Status and Behaviors

When asked whether they had heard of the term "postbiotics," only 13.9% (n=48) of students responded affirmatively, while 86.1% indicated they had never heard of it. A total of 3.8% (n=13) of participants reported having used postbiotic supplements (4.0% n=7 of females vs. 3.5%, n=6 of males). Among those who had used postbiotics, the primary motivations included health-related concerns (30.8%, n=4), recommendations from others (46.2%, n=6), and advertisements (23.0%, n=3). The frequency of postbiotic supplement use was significantly lower among female students compared to males (P<0.001). Table 2 presents a detailed breakdown of students' awareness and consumption behaviors.

Consumption Frequency of Postbiotic-Enhancing Foods

Among foods associated with postbiotic production, yogurt was the most frequently consumed. Specifically, 24.6% (n=85) of students reported daily consumption, and 42.8% (n=148) consumed it twice weekly. Kefir had the lowest consumption rate, with 43.1% (n=149) of students reporting no intake at all. Sauerkraut was most often consumed biweekly 28.6%, (n=99), while 26.6% (n=92) of students reported never consuming pickled vegetables. Overall, pickled vegetables and sauerkraut were consumed more frequently than kefir. No statistically significant differences in food consumption patterns were observed between genders (Table 3).

Postbiotic Knowledge Levels of Students

Postbiotic knowledge was predominantly low

TABLE 2. Consumption Status and Behaviors of Students Regarding Postbiotics Products

	Total n (%)	Female n (%)	Male n (%)	P-value[a]
Have you ever heard of the term "postbiotics"?				
Yes	48 (13.9)	29 (16.8)	19 (11.0)	0.161
No	298 (86.1)	144 (83.2)	154 (89.0)	
Have you ever used a postbiotic supplement before?				
Yes	13 (3.8)	7 (4.0)	6 (3.5)	0.222
No	333 (96.2)	166 (96.0)	167 (96.5)	
If yes, what factors influence your consumption?				
Health Issues	4 (30.8)	4 (57.1)	-	0.097
Advice	6 (46.2)	1 (14.3)	5 (83.3)	
Advertisements	3 (23.0)	2 (28.6)	1 (16.7)	
If you use it, how often do you take postbiotic supplements?				
Frequently	5 (38.5)	2 (28.6)	3 (50.0)	**<0.001**
Rarely	8 (61.5)	5 (71.4)	3 (50.0)	

[a]Fisher's Exact test. Statistically significant P-value is shown in bold.

TABLE 3. Frequency of Consumption of Foods that May Increase Postbiotic Production

	Total n (%)	Male n (%)	Female n (%)	P-value*
Yogurt				
Once a day	85 (24.6)	43 (24.9)	42 (24.3)	X^2=1.384 0.926
2-3 times a day	23 (6.6)	13 (7.5)	10 (5.8)	
Twice a week	148 (42.8)	75 (43.4)	73 (42.2)	
Once every 15 days	65 (18.8)	29 (16.8)	36 (20.8)	
Once a month	20 (5.8)	10 (5.8)	10 (5.8)	
I never consume	5 (1.4)	3 (1.7)	2 (1.2)	
Kefir				
Once a day	26 (7.5)	15 (8.7)	11 (6.4)	X^2=7.200 0.206
2-3 times a day	11 (3.2)	3 (1.7)	8 (4.6)	
Twice a week	24 (6.9)	8 (4.6)	16 (9.2)	
Once every 15 days	88 (25.4)	48 (27.7)	40 (23.1)	
Once a month	48 (13.9)	27 (15.6)	21 (12.1)	
I never consume	149 (43.1)	72 (41.6)	77 (44.5)	
Sauerkraut				
Once a day	31 (9.0)	16 (9.2)	15 (8.7)	X^2=4.052 0.542
2-3 times a day	16 (4.6)	9 (5.2)	7 (4.0)	
Twice a week	92 (26.6)	38 (22.0)	54 (31.2)	
Once every 15 days	99 (28.6)	54 (31.2)	45 (26.0)	
Once a month	61 (17.6)	32 (18.5)	29 (16.8)	
I never consume	47 (13.6)	24 (13.9)	23 (13.3)	
Pickled vegetables				
Once a day	29 (8.4)	15 (8.7)	14 (8.1)	X^2=2.132 0.831
2-3 times a day	25 (7.2)	10 (5.8)	15 (8.7)	
Twice a week	66 (19.1)	33 (19.1)	33 (19.1)	
Once every 15 days	87 (25.1)	43 (24.9)	44 (25.4)	
Once a month	47 (13.6)	27 (15.6)	20 (11.6)	
I never consume	92 (26.6)	45 (26.0)	47 (27.2)	

*Pearson's chi-square test

among participants. According to categorized scale scores, 48.8% (n=169) had poor knowledge, 39.0% (n=135) moderate knowledge, 11.8% good knowledge, and only 0.4% (n=1) very good knowledge. Gender-based comparison revealed that 52.0% (n=90) of female students and 45.7% (n=79) of male students fell into the "poor" knowledge category (Table 4). The mean postbiotic knowledge score for female students was 45.99±7.54, while the mean score for male students was 45.91±7.88. Although the scores were close, the difference was statistically significant (P<0.001), suggesting higher awareness among female participants. Knowledge levels across academic departments showed no statistically significant differences suggesting higher awareness among female participants (Table 5). Knowledge levels across academic departments showed no statistically significant differences (P=0.291) (Table 6).

TABLE 4. Distribution of Participants Postbiotic Knowledge Levels by Gender

Postbiotic knowledge level	Female (n=173)		Male (n=173)		Total (n=346)	
Basic nutrition (Total score: 80)	**n**	**%**	**n**	**%**	**n**	**%**
Poor (<45)	90	52.0	79	45.7	169	48.8
Moderate (45-55)	64	37.0	71	41.0	135	39.0
Good (56-65)	18	10.4	23	13.3	41	11.8
Very Good (>65)	1	0.6	-	-	1	0.4

DISCUSSION

This study investigated postbiotic awareness, consumption behaviors, and knowledge levels among university students from diverse academic disciplines. Despite the growing scientific interest in postbiotics and their potential health benefits, our findings indicate a substantial gap in student knowledge and usage. The International Scientific Association for Probiotics and Prebiotics (ISAPP) defined postbiotics in 2021 as "a preparation of inanimate microorganisms and/or their components that confers a health benefit to the host." This definition proposed by ISAPP is comprehensive enough to allow for the development of postbiotics from various microorganisms and their application to different body sites. Beneficial effects can be evaluated or validated in humans, animals, and other target organs [2].

A study conducted among university students has shown that the average score regarding students' knowledge and perception of probiotics is good [16]. In a study that included healthcare personnel, it was found that 88% of the participants were familiar with the term "probiotic," while 22% were familiar with the term "prebiotic" [17]. In another study conducted with students, 56% of the participants were aware of the definition of probiotics; however, no student was able to answer questions related to the types of bacteria known as probiotics correctly [18]. In a study involving medical students, 57.3% of the students reported that they had never heard the definition of probiotics before [19]. Based on these results, it can be stated that students are familiar with definitions they may have heard in class or through casual conversations, but there is a need to increase their level of knowledge on this topic. In this study, it was found that 86.1% of the students had never heard of the concept of postbiotics (Table 2). It can be said that students' awareness of postbiotics may be low due to the fact that postbiotics is a relatively new concept and postbiotic supplements are not yet widely available on the market [20, 21]. However, this study concluded that only 3.8% of all students use postbiotics (Table 2). Based on this, it is anticipated that the usage rate will increase as awareness of the concept grows. When evaluated alongside the existing probiotic literature, the findings of this study are consistent with previous research demonstrating limited awareness and inconsistent consumption patterns of microbiota-related functional foods among university students.

TABLE 5. Average Postbiotic Knowledge Levels of University Students by Gender

Gender	n	Mean±SD	t	P-value
Male	173	45.91±7.88	0.096	**<0.001**
Female	173	45.99±7.54		

SD, standard deviation. Statistically significant P-value is shown in bold.

TABLE 6. Postbiotic Knowledge Averages of University Students by Academic Departments

Departments	n	Mean±SD	P-value
Recreation	18	47.55±6.98	0.291*
Nursing	42	44.95±7.16	
Recreation management	52	43.90±8.13	
Physical education teaching	37	45.10±6.78	
Coaching education	152	46.92±8.00	
Sports management	45	46.04±6.89	

SD, standard deviation.
*Tukey Test

Similar to studies on probiotics, knowledge levels regarding postbiotics were higher among female students, suggesting that gender-related differences observed in probiotic awareness may also extend to emerging concepts such as postbiotics. These parallels support the notion that postbiotics represent a natural extension of probiotic research rather than an isolated concept.

In this study, no literature was found regarding the frequency of postbiotic consumption, and instead, the studies on the frequency of probiotic food consumption were addressed due to the ability of probiotics to enhance postbiotics. Yalçın *et al*. [22] found in their study that yogurt was the most consumed probiotic food at 90.9%, followed by buttermilk at 59.6% and pickles at 55.6%, while kefir was not preferred. In a study consisting of athletes, the proportion of those who never consumed kefir was found to be 50%, while those who consumed kefir daily were limited to 3.3% [23]. In a study investigating the frequency of probiotic food consumption, the most frequently consumed probiotics among students were yogurt (82.4%), followed by Greek yogurt (55.9%), probiotic-added yogurt (42.4%), and kefir (9.3%) [15]. In this study, the proportion of those who consumed kefir daily was limited to 7.5%, which was lower than the daily consumption rates of other probiotic foods (Table 3). Based on these results, it can be said that individuals prefer products like yogurt and pickles more frequently. When the frequency of consumption of foods that may increase postbiotic production is analyzed, 32.4% of men and 30.1% of women consume yogurt daily. A study conducted by Aydın *et al*. [24], a statistically significant difference was found between the frequency of probiotic food consumption and gender, with women exhibiting a higher frequency of probiotic food consumption compared to men. Another study conducted with adult individuals found that yogurt consumption was significantly higher among female participants compared to male participants [25]. However, in our study, no significant difference was observed based on gender (Table 3). In their research, Sevim *et al*. [23] found that the frequency of pickle consumption was similar to our study, with a consumption rate of 46.6% for 1-2 times a week. The low frequency of sauerkraut consumption has been attributed to limited access, as products like yogurt and ayran are more frequently preferred over sauerkraut in school and dormitory cafeterias. The low consumption frequency of foods other than yogurt may be primarily attributed to university students' lack of awareness regarding the importance of improving health and lifestyle, as well as the fact that a large majority have not heard of the concept of postbiotics.

In a study investigating probiotic knowledge levels, it was observed that 364 participants scored an average of 6.16 out of a maximum of 8 points, indicating that they have a good level of knowledge about probiotics [26]. Özgür and Dinçoğlu [27] stated in their research that university students have some knowledge about probiotics; however, they do not possess sufficient information regarding current developments in the field. However, since the sample group consists of Nutrition and Dietetics students, it can be assumed that they possess knowledge about probiotics. In a study that included medical students, it was found that they had low to moderate levels of knowledge about probiotics and that the students needed more educational programs on the topic. In a study involving 1,126 participants, the level of knowledge about probiotics was investigated, and it was found that 76.4% of the participants had insufficient knowledge, 21.1% had average knowledge, and 2.5% had good knowledge [28]. In our study, the percentage of university students with a very good level of knowledge about postbiotics was found to be 0.4%, while the percentage of those with poor knowledge was 48.8%. Factors contributing to this situation include the fact that the concept of postbiotics is relatively new in our country, insufficient advertising for products containing postbiotics, and low awareness regarding these products (Table 4). When examining the average postbiotic knowledge levels of university students by gender, it was found that females have a higher knowledge level compared to males (Table 5). Horasan *et al*. [29], in their study aimed at determining the probiotic knowledge levels and consumption patterns of university students, found that females consumed probiotic foods at a higher rate, and the difference between genders was statistically significant. Additionally, similar to the findings of this research, there are studies that support the notion that women consume probiotic foods more than men [30, 31]. The reasons for women having a higher level of

knowledge about postbiotics may include their greater attention to health, nutrition, and body image compared to men [32, 33].

Strengths and Limitations

The research is limited to students continuing their education in specific departments of two state universities. Therefore, more comprehensive scientific studies can be conducted on broader sample groups regarding postbiotics. When the strengths of this study are evaluated, the fact that it is one of the first studies to examine the level of postbiotic awareness and knowledge among university students in Turkey and that the subject is relatively new in the literature increases the scientific value of the research. In addition, by selecting participants from Health, Sports and Tourism faculties, comparison of students from different academic fields was made possible and the research results were not limited to health sciences students only.

CONCLUSION

Despite the increasing number of studies demonstrating the positive health effects of postbiotics, it has been found that the knowledge levels and consumption frequencies of postbiotics among students, who are expected to have higher awareness, are not at the anticipated levels. Possible reasons for this situation include the limited promotion and advertising activities regarding postbiotics, a lack of understanding of the health benefits of postbiotics, the infrequent availability of postbiotic products in the market, and uncertainties surrounding their consumption. Therefore, it is recommended to organize training sessions in faculties about postbiotic foods to raise awareness among students. Students who will become future coaches, teachers, and healthcare personnel should first become knowledgeable about this concept, which is recognized for its positive effects on the maintenance and preservation of a healthy lifestyle, and integrate it into their dietary routines. Although there are currently no widespread commercial products or supplements available, postbiotics have the potential to become significant players in the functional food market in the next decade. Therefore, scientists are encouraged to conduct more scientific studies on postbiotics. Therefore, it is anticipated that the consumption frequency of postbiotic foods will increase in light of the scientific studies conducted. Finally, the research is limited to students continuing their education in specific departments of two state universities. Therefore, more comprehensive scientific studies can be conducted on broader sample groups regarding postbiotics.

Ethics Approval and Consent to Participate

This study was approved by the Aksaray University Human Research Ethics Committee (Decision No: 2024/01-41; date: 28.02.20254. All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Written informed consent was obtained from all participants.

Data Availability

All data generated or analyzed during this study are included in this published article. The data that support the findings of this study are available on request from the corresponding author, upon reasonable request.

Authors' Contribution

Study Conception: ANK; Study Design: ANK, YY; Supervision: ANK, YY; Funding: ANK, YY; Materials: ANK; Data Collection and/or Processing: ANK, YY; Statistical Analysis and/or Data Interpretation: ANK; Literature Review: ANK, YY; Manuscript Preparation: ANK, YY; and Critical Review: ANK, YY.

Conflict of Interest

The author(s) disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The author(s) disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgments

The authors have no acknowledgments to declare.

Generative Artificial Intelligence Statement

The author(s) declare that an artificial intelligence-based tool or application was used for grammar editing in the preparation of this article. All content of this work was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

Editor's Note

All statements made in this article are solely those of the authors and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Granato D, Barba FJ, Bursać Kovačević D, Lorenzo JM, Cruz AG, Putnik P. Functional Foods: Product Development, Technological Trends, Efficacy Testing, and Safety. Annu Rev Food Sci Technol. 2020;11:93-118. doi: 10.1146/annurev-food-032519-051708.
2. Vinderola G, Sanders ME, Salminen S. The Concept of Postbiotics. Foods. 2022;11(8):1077. doi: 10.3390/foods11081077.
3. Salminen S, Collado MC, Endo A, et al. The International Scientific Association of Probiotics and Prebiotics (ISAPP) consensus statement on the definition and scope of postbiotics. Nat Rev Gastroenterol Hepatol. 2021;18(9):649-667. doi: 10.1038/s41575-021-00440-6.
4. Ma L, Tu H, Chen T. Postbiotics in Human Health: A Narrative Review. Nutrients. 2023;15(2):291. doi: 10.3390/nu15020291.
5. Li L, Fu H. China's health care system reform: Progress and prospects. Int J Health Plann Manage. 2017;32(3):240-253. doi: 10.1002/hpm.2424.
6. Sharma M, Shukla G. Metabiotics: One Step ahead of Probiotics; an Insight into Mechanisms Involved in Anticancerous Effect in Colorectal Cancer. Front Microbiol. 2016;7:1940. doi: 10.3389/fmicb.2016.01940.
7. Tomasik P, Tomasik P. Probiotics, Non-Dairy Prebiotics and Postbiotics in Nutrition. Appl Sci. 2020;10(4):1470. doi: 10.3390/app10041470.
8. Eraghieh Farahani H, Pourhajibagher M, Asgharzadeh S, Bahador A. Postbiotics: Novel Modulators of Gut Health, Metabolism, and Their Mechanisms of Action. Probiotics Antimicrob Proteins. 2025 Nov 10. doi: 10.1007/s12602-025-10832-8.
9. Li S, Sohouli MH, Li Z. The effect of postbiotics supplementation on obesity and metabolic health: a systematic review and meta-analysis of randomized control trials. Nutr Metab (Lond). 2025;22(1):140. doi: 10.1186/s12986-025-01037-5.
10. Al-Habsi N, Al-Khalili M, Haque SA, Elias M, Olqi NA, Al Uraimi T. Health Benefits of Prebiotics, Probiotics, Synbiotics, and Postbiotics. Nutrients. 2024;16(22):3955. doi: 10.3390/nu16223955.
11. Collado MC, Vinderola G, Salminen S. Postbiotics: facts and open questions. A position paper on the need for a consensus definition. Benef Microbes. 2019;10(7):711-719. doi: 10.3920/BM2019.0015.
12. Wegh CAM, Geerlings SY, Knol J, Roeselers G, Belzer C. Postbiotics and Their Potential Applications in Early Life Nutrition and Beyond. Int J Mol Sci. 2019;20(19):4673. doi: 10.3390/ijms20194673.
13. Cigdem A, Emre N. Nutrition literacy status of university students, influencing factors, and its relationship with healthy nutrition attitudes. Pamukkale Med J. 2025;18(1):105-116. doi: 10.31362/patd.1532810.
14. Elmas N, Bayraktar DZ, Yılmaz B, Çimen F, Özdemir U, Türkel U. Investigating the Relationship Between Nutritional Knowledge and Microbiota Awareness: A Cross-Sectional Study Among University Students. Sağlık Bilimlerinde Değer 2025; 15(3): 367-376. doi: 10.33631/sabd.1600538.
15. Batmaz H. Development of A Nutrition Knowledge Level Scale For Adults and ValidationReliability Study [Master's Thesis]. Istanbul: Marmara University; 2018.
16. Pradito IY, Wardana AA, Waspodo P, Surono IS. Determinants of knowledge and perception of probiotic by Jabodetabek college students. Food Research. 2020;4(5), 1815-1819. doi: 10.26656/fr.2017.4(5).133.
17. Oliver L, Rasmussen H, Gregoire MB, Chen Y. Health care provider's knowledge, perceptions, and use of probiotics and prebiotics. Top Clin Nutr. 2014;29(2), 139-149. doi: 10.1097/01.TIN.0000445898.98017.eb.
18. Al Hossan AA, Syed W, Babelghaith SD, Al Arifi MN. Knowledge, Attitude, and Practice of Probiotics Among Saudi Health Care Students-A Cross-Sectional Study From Saudi University in Riyadh Saudi Arabia. Inquiry. 2024;61:469580231224821. doi: 10.1177/00469580231224821.
19. Chukwu EE, Nwaokorie FO, Yisau JI, Coker AO. Assessment of the Knowledge and Perception of Probiotics Among Medical Science Students and Practitioners in Lagos State. Br J Med Med Res. 2014;5(10)1239-1246. doi: 10.9734/BJMMR/2015/13676.
20. Nataraj BH, Ali SA, Behare PV, Yadav H. Postbiotics-parabiotics: the new horizons in microbial biotherapy and functional foods. Microb Cell Fact. 2020;19(1):168. doi: 10.1186/s12934-020-01426-w.
21. Aggarwal S, Sabharwal V, Kaushik P, Joshi A, Aayushi A, Suri M. Postbiotics: From emerging concept to application. Front Sustain Food Syst. 2022;6,887642. doi: 10.3389/fsufs.2022.887642.
22. Çelik Şahin S, Pasin T, Ankaralı S. The Effect of Probiotic Consumption Status and Quality of Life in Patients with Fibromyalgia. Value Health Sci. 2023;13(3):446-450. doi: 10.33631/sabd.1346342.
23. Sevim Y, Onur Öztürk HN, Ergün M. Knowledge and consumption frequency of probiotics and fermented foods in elite volleyball players - A Pilot Study. J Res Pharm. 2023;27(2):12-21. doi: 10.12991/jrp.2019.00.
24. Aydın İ, Gülsünoğlu Konuşkan Z, Ersoy B. Yetişkin Bireylerin Probiyotik Besinler Hakkında Bilgi Düzeyleri ve

Tüketim Durumlarının Belirlenmesi [The Determination of Probiotic Foods Knowledge Level and Consumption Status of Adults]. Aydın Gastronomy. 2024;8(1):129-141. doi: 10.17932/IAU.GASTRONOMY.2017.016/gastronomy_v08i100. [Article in Turkish]
25. Pehlivan B. Yetişkin Bireylerin Probiyotik Besinleri Tüketim Sıklıklarının ve Bilgi Düzeylerinin Belirlenmesi [Evaluation Frequency of Adults Probiotic Food Consumption and Levels of Knowledge]. Bilimsel Tamamlayıcı Tıp, Regülasyon ve Nöralterapi Dergisi. 2020;14(3):69-79. [Article in Turkish]
26. Mejia WB, Barrion ASA, Abacan SF, Israel KAT. Knowledge and Consumption of probiotic Foods of Selected Students in Laguna, Philippines. EC nutrition. 2019;14(5):452-459.
27. Özgür M, Dinçoğlu AH. Knowledge Levels and Attitudes of Nutrition and Dietetics Department Students Towards Probiotic Products: The Example of Burdur Mehmet Akif Ersoy University: Cross-Sectional Study. J Tradit Complem Med. 2023;6(1):11-23. doi: 10.5336/jtracom.2022-90629.
28. Rajab BS, Jahlan RA, Mobarki AM, et al. Assessment of knowledge, attitudes, and practices regarding microbiota composition and influencing factors among the general population in Jazan province: A cross-sectional study. J Adv Vet Anim Res. 2023;10(4):773-781. doi: 10.5455/javar.2023.j733.
29. Horasan B, Sevinç Ö, Çelikyürek NA. Üniversite Öğrencilerinin Probiyotik Bilgi Düzeyi ve Tüketim Durumlarının Belirlenmesi [Determination of Probiotic Knowledge Level and Consumption Status of University Students]. Avrupa Bilim ve Teknoloji Dergisi. 2021;31(Suppl 1):446-453. doi: 10.31590/ejosat.999946. [Article in Turkish]
30. Yabancı N, Şimşek I. Üniversite Öğrencilerinin Probiyotik Ürün Tüketim Durumları [Status of Probiotic Product Consumption of University Students]. TSK Koruyucu Hekimlik Bülteni. 2007;6(6):449-454. [Article in Turkish]
31. Aydın M, Açıkgöz İ, Şimşek B. Isparta Süleyman Demirel Üniversitesi Öğrencilerinin Probiyotik Ürün Tüketimlerinin ve Probiyotik Kavramının Bilinme Düzeyinin Belirlenmesi [Determination of Probiotic Product Consumption and Probiotic Concept Knowledge Level in Students of Isparta Süleyman Demirel University]. Gıda Teknolojileri Elektronik Dergisi. 2010;5(2):1-6. doi: 10.15237/gida.GD18104. [Article in Turkish]
32. Yücelşengün İ, Kırmızıgül A, Özaydın İ, Yarım H. Tüketicilerin Probiyotik ve Prebiyotik Gıdalara Yönelik Bilgi Düzeyleri ve Tüketim Durumlarının Belirlenmesi: İzmir/Bornova Örneği [Determinatıon of Knowledge Level and Consumptıon Status of Consumers on Probiotıc and Prebiotic Foods: A Sample of Izmır/Bornova]. GIDA. 2020;45(1),103-114. doi: 10.15237/gida.GD19123. [Article in Turkish]
33. Aslan Ş, Kara R, Yaman H. Probiyotik Ürünlerin Tüketim Alışkanlıklarının Belirlenmesi [Determining the Consumption Habits Related to Probiotic Products]. Türk Tarım Gıda Bilim ve Teknoloji Dergisi. 2019;7(6): 861-865. doi: 10.24925/turjaf.v7i6.861-865.2428.

The European Research Journal 2026;12(7):770-779 Research Article

DOI: https://doi.org/10.18621/eurj.1843712 *Public Health*

The Relationship Between Forward Head Posture with Hand Grip Strength and Thoracic Kyphosis Among Young Adults

Yunis Akkaş[1], Ahmet Gökhan Acar[1], Serap Alsancak[1], Senem Güner[1]

[1]*Department of Prosthetics and Orthotics, Ankara University, Faculty Health of Science, Ankara, Türkiye*

Abstract

Objective: Forward head posture (FHP) is a common postural abnormality where the external auditory canal is positioned ahead of the shoulder joint's plumb line. This study aimed to evaluate FHP's impact on thoracic kyphosis, hand grip strength, and pinch grip strength in healthy adults.

Methods: Ninety-three healthy adults (18–25 years) participated, divided into two groups based on craniovertebral angle (CVA): the FHP group (CVA <45°, n=38) and the control group (CVA >45°, n=55). CVA was measured via photogrammetry. The thoracic kyphosis angle was assessed using the Goniometer Pro app. Hand grip strength was measured with a Jamar dynamometer, and pinch grip strength with a Baseline hydraulic pinch gauge.

Results: No significant differences were found between groups in thoracic kyphosis angle, hand grip, or pinch grip strength (P>0.05). CVA showed no significant correlation with these parameters in either group (P>0.05). However, in the FHP group, the thoracic kyphosis angle was significantly correlated with non-dominant hand grip strength (r= -0.324, P=0.047), dominant pinch grip strength (r= -0.350, P=0.031), and non-dominant pinch grip strength (r= -0.394, P=0.014).

Conclusion: Hand grip and pinch grip strength were not affected by FHP in asymptomatic young adults. No relationship was found between CVA and thoracic kyphosis. However, individuals with FHP and thoracic kyphosis angles above 40° may experience altered non-dominant hand function, potentially due to kyphotic changes.

Keywords: Craniovertebral Angle, Forward Head Posture, Hand Grip Strength, Pinch Grip Strength

Many studies have shown that several factors affect hand grip strength, among which a few studies have shown that shoulder and elbow position and head and neck position have an effect [1–3]. In addition, hand strength is related to upper extremity strength, general body strength, and upper extremity function [4]. The hand is important for performing daily life activities that require precise control. Grip strength is the ability of the fingers to grasp objects and is an important factor in hand function performance [5]. The ability to hold small objects between the thumb and index finger is defined as pinch strength [6] Hand grip and pinch strength do not depend solely on the fingers and wrist; they require

Submitted: December 17, 2025 **Accepted:** February 2, 2026 **Published Online:** February 15, 2026

How to cite this article: Akkaş Y, Acar AG, Alsancak S, Güner S. The Relationship Between Forward Head Posture with Hand-Grip Strength and Thoracic Kyphosis Among Young Adults. *Eur Res J.* 2026;12(7):770-779. doi: 10.18621/eurj.1843712

Corresponding author: Yunis Akkaş, Ph.D., Phone: +90 312 319 50 18, E-mail: yakkas@ankara.edu.tr

Available Online at https://www.eurj.org.tr

proper functioning of the forearm, pre-scapular, and shoulder muscles [7]. Scapular position affects upper extremity and hand function. Upper extremity function may be affected by a hyperkyphotic posture, which involves scapular protraction and scapulohumeral rhythm dissociation.

Forward head posture (FHP), defined as any alignment in which the external auditory canal is positioned in front of the plumb line across the shoulder joint, can develop when using a smartphone, carrying a backpack, using a computer, overusing the shoulders, or maintaining poor posture habits [8–12]. FHP has also been associated with cervical spine hyperextension, forward head tilt, and shortened levator scapulae, sternocleidomastoid, upper trapezius, and posterior cervical spine muscles. In people with FHP, when performing arm lifting activities, a decrease in serratus anterior activity causes the scapula to bend forward, and at the same time, winging of the scapula occurs. With winging of the scapula, the upper thoracic slope greatly increases in the transverse plane, and thoracic kyphosis can occur, with an increase in the anteroposterior diameter of the rib cage. The posture of the thoracic spine also affects scapular and glenohumeral kinematics. Increased flexion of the thoracic spine can lead to decreased range of motion at the glenohumeral joint [12, 13]. The relationship between increased FHP and thoracic kyphosis has been explored in previous studie [14, 15].

Few studies have investigated the relationship between excessive smartphone use and hand grip strength or pinch grip strength. One study reported that the long-term use of smartphones is associated with poorer hand grip and pinching, showing that the duration of smartphone use, a factor other than age, may contribute to reduced hand muscle strength. [16]. A study comparing the hand grip and pinch strength of children with high-frequency and low-frequency smartphone use found that the high-frequency users had decreased hand and pinch strength [17].

Long-term smartphone use commonly leads to symptoms such as FHP, decreased cervical range of motion, and pain around the neck [18, 19]. However, there are limited studies examining the relationship between FHP and thoracic kyphosis angle, hand grip, and pinch grip strength. Therefore, this study was conducted to evaluate the relationships and differences between hand and finger muscle strength, thoracic kyphosis angle, and FHP in individuals with and without common FHP symptoms. The study hypothesized that, in individuals with FHP syndrome, there would be a detectable relationship between the FHP angle and the thoracic kyphosis angle, which may affect the upper extremity kinematics and, in turn, negatively affect hand grip strength and pinch grip strength.

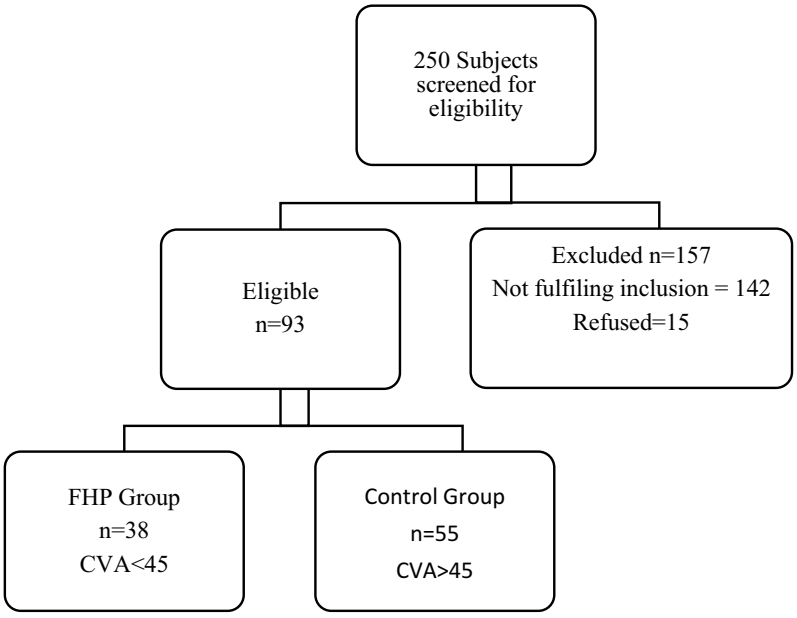

FIGURE 1. The participants' flow chart

METHODS

Participants

This study, conducted at the Prosthetics and Orthotics Department of Ankara University from September 2023 to January 2024, evaluated the FHP degree, thoracic kyphosis degree, and hand and pinch grip strength in healthy adults. A random sample of two hundred and fifty participants was composed of undergraduate students of Ankara University Faculty of Health Sciences. Ninety-three students between the ages of 18 and 25 were recruited for this study. A priori sample size calculation was performed using G*Power software (version 3.1.9.7; Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany). Based on a medium-to-large effect size (Cohen's d = 0.6), an alpha level of 0.05, and a power of 80% for a two-tailed independent samples t-test, the minimum required total sample size was calculated as 90 participants. Therefore, our sample of 93 participants was considered sufficient to provide adequate statistical power. The participants included in the study had at least 6 months of experience using a smartphone, spent 2 h or more per day on a smartphone, and achieved minimum smartphone texting and computer typing speeds of between 15 and 30 words per minute, respectively. Smartphone users in both groups were excluded from the study if they had a history of traumatic injury, neck pain, other medical conditions or surgical interventions involving the spine and upper extremities, an orthopedic or neurological disorder, mental illness, or a chronic disorder affecting the musculoskeletal system, such as rheumatoid arthritis, osteoarthritis, or another connective tissue disease and if they were athletes (Figure 1). Upon acceptance of participation, informed consent was obtained for signature, and then the evaluation form was filled out by the researcher. FHP and thoracic angle measurements were made by the same physiotherapist.

Each participant provided written informed consent according to the Declaration of Helsinki, and the study procedure was explained and demonstrated to them. This study was approved by the Ankara University Faculty 0f Medicine Human Research Ethics Committee.

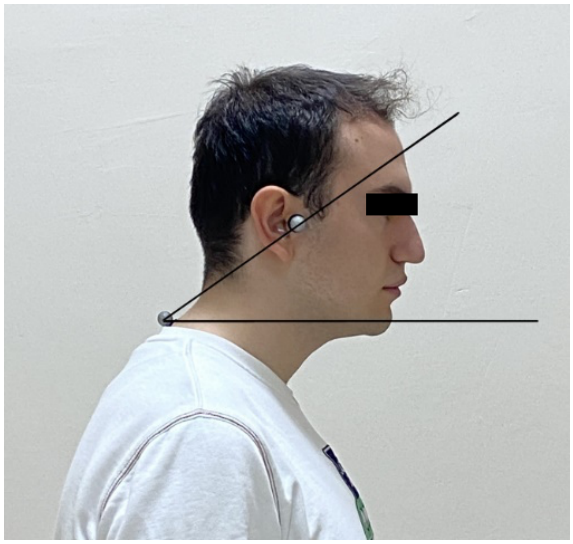

FIGURE 2. Measurement of craniovertebral angle.

Procedure for Assessment

Measurement of Craniovertebral Angle

In this cross-sectional study, we compared a group of 38 young adults over the age of 18 years with FHP <45° to a group of 55 matched individuals with FHP >45° who had normal FHP. For the assessment of FHP, the craniovertebral angle (CVA) was measured using the photogrammetric method. The tragus and 7th cervical vertebra (C7) points were taken as a reference based on photographs of the individual's craniocervical angle in the sagittal plane. The angle between a line drawn from the tragus and a line drawn vertically from C7 was recorded as the anterior tilt angle. Photographs were taken with the individual standing in a comfortable position with feet shoulder-width apart and head facing forward. Subjects were asked to stand 1.5 m from the postural board, and photographs were taken with the digital camera on a tripod. All photographs were analyzed using the ImageJ program [20–23]. An angle of less than 45.5° indicates anterior tilt to the head (Figure 2).

Assessment of Hand and Pinch Grip Strength in Dominant and Non-Dominant Hands

Measurements were taken from the dominant and non-dominant hands in both groups. The dominant hand was defined as the one preferred for daily activities such as writing, eating and handling heavy objects. Grip strength was measured using a Jamar hand dynamometer (Jamar, Los Angeles, CA, USA), and

pinch strength was measured using a Baseline hydraulic pinch gauge (Fabrication Enterprises, Inc.). The Jamar hand dynamometer, which is recommended by the American Society of Hand Therapists as a standard tool for measuring grip strength, is small and portable. It was professionally calibrated before use according to the manufacturer's instructions. The dial indicates force in both kilograms and pounds, marked in 2 kg or 5 lb increments, allowing evaluation to the nearest 1 kg or 2.5 lb. The Jamar dynamometer is adjustable to accommodate hands of various sizes [24, 25]. Each subject was instructed to sit in a straight-backed chair. Hand muscle strength was measured with feet flat on the ground, shoulders in adduction, elbow joint in 90° flexion, and forearm in neutral position, with the wrist joint at 0° and 30° extension and 0° and 15° ulnar deviation [26]. Three measurements of hand grip strength were registered for each hand, with a 1 min rest between measurements to avoid fatigue, then the average value was recorded [27, 28]. Pinch grip strength was measured as tip pinch (thumb tip to index tip, also known as 2-point pinch) [29]. The average of 3 trials was recorded for final analysis.

Measurement of Thoracic Kyphosis

The researcher explained to the evaluators how to take measurements and find skeletal landmarks. T1–T12 signs were labeled as follows: First, the C7 spinous process was located by palpation. This vertebra has the largest protrusion in the neck region when the head is tilted forward. Then, the T1 spinous process was identified by touching the bottom of C7. The subject was asked to flex the trunk slightly forward, and the T2–T12 spinous processes were located by palpation [30].

A Samsung Galaxy Note9 smartphone with a 6.3-inch screen with the Goniometer Pro Android application was used [31]. The validity and reliability of the Goniometer Pro application for measuring spinal curvature have been well-established in recent literature. Shahri and Hesar [32] reported a strong correlation (r = 0.81) and a high intraclass correlation coefficient (ICC = 0.89) between the Goniometer Pro measurements and the radiographic Cobb angle, which is considered the gold standard for thoracic kyphosis assessment. Similarly, Pakeloğlu *et al.* [33] demonstrated that the application possesses excellent intra-rater reliability (ICC = 0.90) and high inter-rater reliability (ICC = 0.87) specifically for measuring the thoracic kyphosis angle in university students. Furthermore, previous studies evaluating the electromechanical properties of smartphone-based goniometers have confirmed their accuracy, with measurement errors of less than 2° compared to universal goniometers [34]. First, we ran the Goniometer Pro application and placed the middle of the lower edge of the smartphone on the marked point of the spinous process (T1), and when the angle was shown, we tapped the purple button on the screen. We then repeated the process for T12 and tapped the green circle on the phone screen to show the scores. The lowest score was the kyphosis angle. If the calculated angle was equal to or greater than 40°, the participant was assigned to the hyperkyphosis group [35].

Statistical Analysis

It was performed using SPSS Statistics (version 26; IBM, Chicago, IL, USA). The Kolmogorov–Smirnov test was used to assess the normality of the data distribution. Descriptive statistics and t-tests were used for the comparisons of mean age, height, weight, and body mass index. To compare thoracic kyphosis, hand grip, and pinch strength between groups, we used independent sample t-tests. The Spearman correlation coefficient was used to determine the correlation between CVA and thoracic kyphosis and grip and pinch strength. The level of significance for all statistical tests was set at $P<0.05$.

RESULTS

General Characteristics of the Participants

There were 93 participants in this study; Table 1 shows their demographic features, CVA, thoracic kyphosis angle, grip strength, and pinch strength. The FHP group (n=38) had a CVA of less than 45°, and the control group (n=55) had a CVA of more than 45°. There was a significant difference in CVA between the groups ($P<0.001$): the average CVA was 42.87±2.4° in the FHP group and 50.81±3.6° in the control group. There were no significant differences between groups in terms of mean age and height. Average mean body mass index and weight were significantly different

TABLE 1. Descriptive Characteristics of the Study Groups

	FHP group CVA<45 degrees (n=38)	Control group CVA>45 degrees (n=55)	P-value
Age (years)	21.15±2.3	21.18±4	0.970
Height (cm)	168.81±10.6	168.86±10.1	0.980
Weight (kg)	70.26±15.5	62.2±12.7	0.008
BMI (kg/m^2)	24.42±3.4	21.67±3	**<0.001**
CVA (degrees)	42.87±2.4	50.81±3.6	**<0.001**
Thoracic kyphosis angle (degrees)	43.88±9	43.33±6.6	0.790
Dominant hand grip strength	33.61±10.6	31.57±10.7	0.298
Non-dominant hand grip strength	32.19±11	29.35±10.7	0.147
Dominant pinch grip strength	20.57±5.4	19.15±4.9	0.275
Non-dominant pinch grip strength	19.42±4.4	18.44±4.4	0.255

Data are shown as mean±standard deviation. BMI, body mass index; CVA, craniovertebral angle; FHP, forward head posture. Statistically significant P-values are shown in bold.

between groups (P<0.05): the average BMI (kg/m^2) was 24.42±3.4 in the FHP group and 21.67±3 in the control group. No significant differences in the mean thoracic kyphosis angle, hand grip, and pinch grip strength of both hands were found between groups (P>0.05).

Correlation Analysis

As shown in Table 2, no significant correlations between CVA and thoracic kyphosis angle, hand grip, and pinch grip strength were found in both groups in this study. Table 3 shows the correlations between thoracic kyphosis degree and CVA, hand grip, and pinch grip strength. In the FHP group (CVA <45°), significant correlations were found between thoracic kyphosis degree and non-dominant hand grip strength (r= -0.324, P=0.047), dominant hand pinch strength (r= -0.350, P=0.031), and non-dominant hand pinch strength (r= -0.394, P=0.014), but there was no significant correlation with dominant hand grip strength. The thoracic kyphosis angle showed a weak and negative correlation with non-dominant hand grip strength, dominant hand pinch strength, and non-dominant hand pinch strength. Table 3 shows the correlations between grip strength and pinch strength; significant differences were found between dominant and non-dominant hand grip strength and dominant and non-dominant hand pinch strength in the two groups. Hand grip strength showed positive correlation with pinch grip strength.

TABLE 2. Correlation Between CVA Degree on Thoracic Kyphosis Angle, Hand, and Tip Pinch Grip Strength

CVA degree	Thoracic kyphosis angle, degrees	Dominant hand grip strength	Non-dominant hand grip strength	Dominant pinch grip strength	Non-dominant pinch grip strength
FHA Group	r=-0.066 P=0.694	r=0.026 P=0.875	r=0.058 P=0.728	r=-0.049 P=0.769	r=-0.063 P=0.706
Control Group	r=-0.156 P=0.254	r=-.140 P=0.308	r=-.117 P=0.384	r=-.004 P=0.980	r=.004 P=0.975

FHA group, CVA<45 degrees; Control group, CVA>45 degrees; CVA, craniovertebral angle

TABLE 3. Correlation Between Thoracic Kyphosis Angle on FHP Degree, Hand, and Tip Pinch Grip Strength

Thoracic kyphosis angle	FHP degree	Dominant hand grip strength	Non-dominant hand grip strength	Dominant pinch grip strength	Non-dominant pinch grip strength
FHP group	r=-0.066	r=-0.263	r=-0.324	r=-0.350	r=0-.394
	P=0.694	P=0.111	**P=0.047**	**P=0.031**	**P=0.014**
Control group	r=-0.156	r=0.043	r=0.044	r=0.039	r=0.012
	P=0.254	P=0.758	P=0.747	P=0.778	P=0.930

FHA group, CVA<45 degrees; Control group, CVA>45 degrees; CVA, craniovertebral angle; FHP, forward head posture. Statistically significant P-values are shown in bold.

DISCUSSION

The purpose of this study was to evaluate the correlations between thoracic kyphosis, dominant and non-dominant hand grip strength, and dominant and non-dominant hand pinch grip strength in young smartphone users with CVA <45° and CVA >45°. The results show that the thoracic kyphosis angle, hand grip strength, and pinch grip strength were not associated with CVA. There was no difference in thoracic kyphosis angle, hand grip strength, or pinch grip strength between the two groups. However, the thoracic kyphotic angle was correlated with the grip strength and pinch strength of the non-dominant hand in the FHP group, and this may be related to head and neck rotational position [31, 36].

FHP is defined as a CVA of less than 45° and manifests as forward protrusion of the head in the sagittal plane and extension in the upper cervical region. FHP can cause neck pain, changes in postural control, and decreased balance ability. Any change in cervical lordosis can biomechanically lead to postural changes in the thoracic and lumbar spine. We did not find any correlation between CVA and thoracic kyphosis angle. The current results are supported by the findings of Talati *et al*., who did not find any significant correlation between FHP, thoracic kyphosis, and lumbar lordosis in healthy adults aged 18–35 years [37].

Lee *et al*. [38] evaluated the use of electromyography to record the muscle activity of the upper trapezius, extensor pollicis longus, and abductor pollicis during smartphone use and observed higher activity in these muscles during one-handed smartphone use than during two-handed use. The study noted that increased muscle activity may compensate for instability in the shoulder and wrist during single-handed device handling. We did not find a correlation between CVA and hand grip or pinch grip strength, but did find a correlation between thoracic kyphosis angle and non-dominant hand grip strength and both dominant and non-dominant hand pinch grip strength in the FHP group. Because the people evaluated in our study were using a smartphone with their dominant hand and there was an increase in upper extremity muscle activity on the dominant side, a relationship with the dominant hand may not have been detectable. In our study, we found a relationship between the thoracic kyphosis angle and the pinch strength on the non-dominant hand side, which may align with the findings of Lee *et al*. Because of the increased muscle strength and compensation on the dominant side, there may be no relationship between CVA and thoracic kyphosis angle.

Several factors may explain why our hypothesis regarding the correlation between CVA, thoracic kyphosis, and grip strength was not supported. First, the participants in this study were young, asymptomatic adults (18–25 years). In this age group, postural deviations such as FHP are often 'functional' or flexible in nature, originating from soft tissue habits rather than fixed structural bony deformities [37]. Therefore, a forward head position may not mechanically force the thoracic spine into a rigid hyperkyphotic curve as strictly as it would in older populations with degenerative changes. Second, young adults possess high biomechanical compensatory capabilities. As suggested by Lee *et al*., increased muscle activity in the upper extremities during smartphone use may compensate for the

glenohumeral instability caused by poor posture, thereby maintaining grip strength levels despite the altered alignment [38]. Consequently, the kinematic chain effect—where FHP leads to kyphosis and subsequently reduces grip strength—may not yet be fully established in this early stage of postural misalignment.

A study by Mosaad *et al*. [39] reported that hand grip strength was not affected by forward head tilt and rounded shoulder posture or by rounded shoulder posture alone in asymptomatic young adults, and CVA was not associated with an inverse effect on hand grip strength. Mosaad *et al*.'s [39] study supports the results of this study, with no correlation found between CVA and hand grip and pinch grip strength. In the Mosaad *et al*.'s [39] study, the average CVA was 47.03±2.42° in the group with forward head tilt and rounded shoulder posture, whereas in our study, the CVA was 42.87±2.4°, and even if the angle was smaller, there was no correlation with grip or pinch strength. The lack of correlation between CVA and grip strength was consistent with the findings of Mosa and Fayaz [40], who concluded that there was no significant relationship between FHP severity and upper extremity anthropometry, including total upper extremity, forearm, and hand length, in addition to mid-arm circumference. Alshahrani *et al*. [41] reported that smartphone addiction negatively affected neck flexor endurance but not grip or pinch strength in healthy college students. The current results are not supported by the findings of Bashir *et al*. [42], who found an increase in FHP; increased disabilities of the arm, shoulder, and hand; and a probable decrease in the grip strength of the dominant hand in smartphone users.

The discrepancy between the findings of Mosaad *et al*. [39] and Bashir *et al*. [42], may be attributed to differences in the study populations and the severity of symptoms. Mosaad *et al*.'s [39] study, similar to ours, evaluated asymptomatic young adults where postural deviations might be in a flexible stage without causing significant biomechanical disadvantages in the upper extremity [39]. In contrast, Bashir *et al*.'s [42] study likely included participants with more pronounced musculoskeletal symptoms or different patterns of smartphone usage intensity, which could lead to detectable deficits in grip strength. Therefore, the impact of FHP on hand function appears to be context-dependent, becoming more evident when accompanied by pain or chronic disability rather than in asymptomatic structural variations.

The correct position of the scapula increases upper extremity movement efficiency and its stability affects hand functions. In kyphotic individuals, the scapula is in a protracted position, disrupting the scapulohumeral rhythm. Limited scapular and chest movement also reduces the shoulder joint's range of motion [43]. The kyphotic spinal structure can cause muscle imbalance in the upper extremity kinetic chain, affecting the pectoralis minor, pectoralis major, rhomboids, and serratus anterior. This imbalance can result in reduced grip and pinch strength [44].

In our study, there was no difference between the thoracic kyphosis angle in the FHP group (43.88±9°) and the control group (43.33±6.6°). Both were over 40°, and there was a slightly kyphotic structure. In adults, the kyphosis angle can vary between ~35° and 37° according to different investigators, although those studies were conducted in heterogeneous populations [45]. Boseker *et al*. [46] reported that normal kyphosis ranged between 20° and 50° using±two standard deviations.

In our study, a negative relationship between thoracic kyphosis and non-dominant hand grip strength and dominant and non-dominant hand pinch strength was found in the FHP group. The correlation between thoracic kyphosis and non-dominant hand and pinch grip strength may be due to muscle activation on the dominant side and increased activation of the extensor pollicis longus and abductor pollicis muscles, especially in one-sided smartphone users. Second, there may be an effect depending on the head and neck position. Kumar *et al*. [3] concluded that head–neck position can influence grip strength and noted that, for right-hand-dominant people, head–neck rotation to the left may have a greater influence than a neutral position or rotation to the right. Fercho *et al*. [47] identified the region that makes greatest contribution to forward flexion along the cervical parameters during various tasks involving smartphone use, and indicated that the greatest contributor to head flexion is the C0–C1 joint and involuntary rotation of the cervical spine toward symmetry when texting. Involuntary rotation in the cervical region can affect grip and pinch strength in individuals with FHP. According to these studies, the functional effects of the muscles that rotate the cervical region occur in

individuals with FHP. A study by Xie *et al*. [48] found a significant increase in right-side cervical flexion angle and greater postural changes in cervical rotation when texting on a smartphone. While texting with two hands was associated with increased cervical flexion, texting with one hand was associated with asymmetric movement.

Strengths and Limitations

This study provides valuable insights into the relationship between sagittal plane postural alignments and distal extremity functions, specifically focusing on FHP and grip strength. A key strength of the research is the objective measurement of the craniovertebral angle (CVA) and thoracic kyphosis, which allowed for a precise categorization of the subjects. Furthermore, by focusing on a specific, young, and asymptomatic population, the study minimizes the confounding effects of age-related degenerative changes or pre-existing musculoskeletal pathologies, offering a clearer view of the kinetic chain connection in a modern, smartphone-using demographic.

Despite these strengths, the current study has some limitations. First, only young and healthy individuals were included; therefore, the results cannot be generalized to the broader population or clinical groups, though they serve as a solid basis for future research. Second, the cross-sectional design of the study limits our ability to establish causal relationships. Future research should employ longitudinal designs to determine if long-term, asymptomatic FHP eventually leads to structural changes and functional deficits in grip strength as individuals age. Additionally, incorporating electromyographic analysis of the cervical and scapulothoracic muscles would provide deeper insights into the compensatory muscle activation patterns. Finally, interventional studies examining whether corrective postural exercises for FHP result in improvements in distal hand function would further clarify the clinical relevance of the kinetic chain connection.

CONCLUSION

Based on the data analysis, it can be concluded that there was no correlation between thoracic kyphosis and hand grip and pinch grip strength in individuals with FHP compared to those without FHP. In our findings, no relationship between FHP and thoracic kyphotic angle was detected, but a negative correlation between thoracic kyphosis and non-dominant hand grip strength and pinch grip strength in the FHP group was detected. In the control group, no correlation between thoracic kyphosis and non-dominant hand grip and pinch strength was found. In individuals with FHP with a thoracic kyphosis angle greater than 40°, non-dominant hand functions may be affected. FHP may affect upper extremity kinetics and, upon evaluating joint movement and muscle strength in the upper extremity and cervical region of individuals with FHP in the clinic, recommendations and orthotic approaches can be suggested to keep the head and neck position neutral.

Ethics Approval and Consent to Participate

Ethical approval was granted by the Ankara University Faculty of Medicine Human Research Ethics Committee (Decision No: İ03-193-24; Date: 21.03.2024). This study was conducted in accordance with the ethical standards of the institutional research committee and the 1964 Helsinki Declaration and its later amendments. All participants provided written informed consent prior to their inclusion in the study.

Data Availability

All data generated or analyzed during this study are included in this published article. The data that support the findings of this study are available on request from the corresponding author, upon reasonable request.

Authors' Contribution

Study Conception: SG, SA; Study Design: SG, SA; Supervision: SG, SA; Funding: N/A; Materials: N/A; Data Collection and/or Processing: SG, SA, YA; Statistical Analysis and/or Data Interpretation: SG, SA, AGA; Literature Review: SG, SA; Writer: SG, SA, YA; and Critical Review: SG, SA, YA, AGA.

Conflict of Interest

The author(s) disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The author(s) disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgments

The authors would like to thank the participants for their valuable contribution to this study.

Generative Artificial Intelligence Statement

The author(s) declare that no artificial intelligence-based tools or applications were used during the preparation process of this manuscript. The all content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

Editor's Note

All statements made in this article are solely those of the authors and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Zafar H, Alghadir A, Anwer S. Effects of Head-Neck Positions on the Hand Grip Strength in Healthy Young Adults: A Cross-Sectional Study. Biomed Res Int. 2018;2018:7384928. doi: 10.1155/2018/7384928.
2. Lee JH, Yoo WG, Lee KS. Effects of head-neck rotation and kinesio taping of the flexor muscles on dominant-hand grip strength. J Phys Ther Sci. 2010;22(3):285-289. doi: 10.1589/jpts.22.285.
3. Kumar NS, Daniel CR, Hilda M, Dharmarajan R. Grip strength: influence of head-neck position in normal subjects. J Neurol Res. 2012;2(3):93-98. doi: 10.4021/jnr117w.
4. Balogun JA, Akomolafe CT, Amusa LO. Grip strength: effects of testing posture and elbow position. Arch Phys Med Rehabil. 1991;72(5):280-283.
5. Bonfiglioli C, De Berti G, Nichelli P, Nicoletti R, Castiello U. Kinematic analysis of the reach to grasp movement in Parkinson's and Huntington's disease subjects. Neuropsychologia. 1998;36(11):1203-1208. doi: 10.1016/s0028-3932(97)00171-1.
6. Nataraj R, Evans PJ, Seitz WH Jr, Li ZM. Effects of carpal tunnel syndrome on reach-to-pinch performance. PLoS One. 2014;9(3):e92063. doi: 10.1371/journal.pone.0092063.
7. Park JY, Lee SH, Oh JH, Kim HK. Scapular dyskinesis. J Korean Shoulder Elbow Soc. 2009;12(2):271-277. doi: 10.5397/CiSE.2009.12.2.271.
8. Jung SI, Lee NK, Kang KW, Kim K, Lee DY. The effect of smartphone usage time on posture and respiratory function. J Phys Ther Sci. 2016;28(1):186-189. doi: 10.1589/jpts.28.186.
9. Straker LM, O'Sullivan PB, Smith A, Perry M. Computer use and habitual spinal posture in Australian adolescents. Public Health Rep. 2007;122(5):634-643. doi: 10.1177/003335490712200511.
10. Chansirinukor W, Wilson D, Grimmer K, Dansie B. Effects of backpacks on students: measurement of cervical and shoulder posture. Aust J Physiother. 2001;47(2):110-116. doi: 10.1016/s0004-9514(14)60302-0.
11. Weon JH, Oh JS, Cynn HS, Kim YW, Kwon OY, Yi CH. Influence of forward head posture on scapular upward rotators during isometric shoulder flexion. J Bodyw Mov Ther. 2010;14(4):367-374. doi: 10.1016/j.jbmt.2009.06.006.
12. Singla D, Veqar Z. Association Between Forward Head, Rounded Shoulders, and Increased Thoracic Kyphosis: A Review of the Literature. J Chiropr Med. 2017;16(3):220-229. doi: 10.1016/j.jcm.2017.03.004.
13. Lynch SS, Thigpen CA, Mihalik JP, Prentice WE, Padua D. The effects of an exercise intervention on forward head and rounded shoulder postures in elite swimmers. Br J Sports Med. 2010;44(5):376-381. doi: 10.1136/bjsm.2009.066837.
14. Lau KT, Cheung KY, Chan KB, Chan MH, Lo KY, Chiu TT. Relationships between sagittal postures of thoracic and cervical spine, presence of neck pain, neck pain severity and disability. Man Ther. 2010;15(5):457-62. doi: 10.1016/j.math.2010.03.009.
15. Joshi S, Balthillaya G, Neelapala YVR. Thoracic Posture and Mobility in Mechanical Neck Pain Population: A Review of the Literature. Asian Spine J. 2019;13(5):849-860. doi: 10.31616/asj.2018.0302.
16. Osailan A. The relationship between smartphone usage duration (using smartphone's ability to monitor screen time) with hand-grip and pinch-grip strength among young people: an observational study. BMC Musculoskelet Disord. 2021;22(1):186. doi: 10.1186/s12891-021-04054-6.
17. Radwan NL, Ibrahim MM, Mahmoud WSE. Evaluating hand performance and strength in children with high rates of smartphone usage: an observational study. J Phys Ther Sci. 2020;32(1):65-71. doi: 10.1589/jpts.32.65.
18. Alzaid AN, Alshadoukhi O, Alnasian A. The Prevalence of Neck Pain and the Relationship between Prolonged Use of Electronic Devices and Neck Pain in a Saudi Arabia: Cross - Sectional Study in Saudi Arabia. Egypt J Hosp Med. 2018;70(12):1992–1999. doi: 10.12816/0044856.
19. Yoon W, Han H, Choi S, Shin G. Neck muscle activation and head kinematics when using a smartphone while walking. Proc Hum Factors Ergon Soc Annu Meet. 2019;63(1):957–961. doi: 10.1177/1071181319631184.
20. Ruivo RM, Pezarat-Correia P, Carita AI. Intrarater and interrater reliability of photographic measurement of upper-body standing posture of adolescents. J Manipulative Physiol Ther. 2015;38(1):74-80. doi: 10.1016/j.jmpt.2014.10.009.
21. do Rosário JL. Photographic analysis of human posture: a literature review. J Bodyw Mov Ther. 2014;18(1):56-61. doi: 10.1016/j.jbmt.2013.05.008.
22. Roberts HC, Denison HJ, Martin HJ, et al. A review of the measurement of grip strength in clinical and epidemiological studies: towards a standardised approach. Age Ageing.

2011;40(4):423-429. doi: 10.1093/ageing/afr051.
23. Rasband WS. ImageJ. Bethesda (MD): U.S. National Institutes of Health; 2007. Available from: https://imagej.nih.gov/ij/
24. Cooper C. Fundamentals of hand therapy: clinical reasoning and treatment guidelines for common diagnoses of the upper extremity. St. Louis (MO): Elsevier Health Sciences; 2013.
25. Mathiowetz V, Wiemer DM, Federman SM. Grip and pinch strength: norms for 6- to 19-year-olds. Am J Occup Ther. 1986;40(10):705-711. doi: 10.5014/ajot.40.10.705.
26. Xu ZY, Gao DF, Xu K, Zhou ZQ, Guo YK. The Effect of Posture on Maximum Grip Strength Measurements. J Clin Densitom. 2021;24(4):638-644. doi: 10.1016/j.jocd.2021.01.005.
27. Horsley I, Herrington L, Hoyle R, Prescott E, Bellamy N. Do changes in hand grip strength correlate with shoulder rotator cuff function? Shoulder Elbow. 2016;8(2):124-129. doi: 10.1177/1758573215626103.
28. Lam NW, Goh HT, Kamaruzzaman SB, Chin AV, Poi PJ, Tan MP. Normative data for hand grip strength and key pinch strength, stratified by age and gender for a multiethnic Asian population. Singapore Med J. 2016;57(10):578-584. doi: 10.11622/smedj.2015164.
29. Walukonis K, Beasley J, Boerema R, Powers J, Anderson K. The impact of finger position on pinch strength. Hand Ther. 2018;23(2):70-76. doi: 10.1177/1758998317752966.
30. Van Sint Jan S. Color atlas of skeletal landmark definitions: guidelines for reproducible manual and virtual palpations. Edinburgh: Churchill Livingstone Elsevier; 2007.
31. Faramarzi Kohneh Shahri Y, Ghani Zadeh Hesar N. Validity and reliability of smartphone-based Goniometer-Pro app for measuring the thoracic kyphosis. Musculoskelet Sci Pract. 2020;49:102216. doi: 10.1016/j.msksp.2020.102216.
32. Shahri YFK, Hesar NGZ. Validity and reliability of smartphone-based Goniometer-Pro app for measuring the thoracic kyphosis. Musculoskelet Sci Pract. 2020;49:102216. doi: 10.1016/j.msksp.2020.102216.
33. Pakeloğlu AC, Koç M, Yılmaz AS, Bayar M, Bayar K. Comparing the reliability of the Goniometer Pro application and flexicurve for measuring thoracic kyphosis: a cross-sectional study. Int J Ther Rehabil. 2023;30(11):1–9. doi: 10.12968/ijtr.2023.0013.
34. Wellmon RH, Gulick DT, Paterson ML, Gulick CN. Validity and reliability of 2 goniometric mobile apps: device, application, and examiner factors. J Sport Rehabil. 2016;25(4):371–379. doi: 10.1123/jsr.2015-0041.
35. Anbarian M, Mokhtari M, Zareie P, Yalfani A. A Comparison of Postural Control Characteristics Between Subjects with Kyphosis and Controls. Avicenna J Clin Med. 2010;16(4):53-60.
36. Amin DI, Hawari MZ, Hassan HES, Elhafez HM. Effect of sex and neck positions on hand grip strength in healthy normal adults: a cross-sectional observational study. Bull Fac Phys Ther. 2016;21:42-47. doi: 10.4103/1110-6611.188028.
37. Talati D, Varadhrajulu G, Malwade M. The effect of forward head posture on spinal curvatures in healthy subjects. Asian Pac J Health Sci. 2018;5(1):60-63. doi: 10.21276/apjhs.2018.5.1.13.
38. Lee M, Hong Y, Lee S, et al. The effects of smartphone use on upper extremity muscle activity and pain threshold. J Phys Ther Sci. 2015;27(6):1743-1745. doi: 10.1589/jpts.27.1743.
39. Mosaad DM, Abdel-Aziem AA, Mohamed GI, Abd-Elaty EA, Mohammed KS. Effect of forward head and rounded shoulder posture on hand grip strength in asymptomatic young adults: a cross-sectional study. Bull Fac Phys Ther. 2020;25(1):1-8. doi: 10.1186/s43161-020-00001-z.
40. Mosa DM, Fayaz NA. Association Between Forward Head Posture and Upper Limb Anthropometry in Healthy Adults. Med J Cairo Univ. 2017;85(7):2455-2460.
41. Alshahrani A, Samy Abdrabo M, Aly SM, et al. Effect of Smartphone Usage on Neck Muscle Endurance, Hand Grip and Pinch Strength among Healthy College Students: A Cross-Sectional Study. Int J Environ Res Public Health. 2021;18(12):6290. doi: 10.3390/ijerph18126290.
42. Bashir U, Noor R, Shoukat H, Ali ML, Javed MT, Hassan Z. Correlation of mobile phone usage on grip strength, disabilities and posture in young adults. Rehabil J. 2023;7(1):495-498. doi: 10.52567/trj.v7i01.210.
43. Ardakani MK, Fard ZS, Amirizadeh F, Naderifar H. Effect of Thoracic Hyper-Kyphosis Posture on Upper Extremity Function of Female Students. J Rehabil Sci Res. 2022;9(1):30-35. doi: 10.30476/jrsr.2021.93346.1232.
44. Sahrmann S. Diagnosis and treatment of movement impairment syndromes. 1st ed. St. Louis (MO): Mosby/Elsevier; 2002.
45. Bernhardt M, Bridwell KH. Segmental analysis of the sagittal plane alignment of the normal thoracic and lumbar spines and thoracolumbar junction. Spine (Phila Pa 1976). 1989;14(7):717-721. doi: 10.1097/00007632-198907000-00012.
46. Boseker EH, Moe JH, Winter RB, Koop SE. Determination of "normal" thoracic kyphosis: a roentgenographic study of 121 "normal" children. J Pediatr Orthop. 2000;20(6):796-798. doi: 10.1097/00004694-200011000-00019.
47. Fercho J, Krakowiak M, Yuser R, et al. Kinematic analysis of the forward head posture associated with smartphone use. Symmetry (Basel). 2023;15(3):667. doi: 10.3390/sym15030667.
48. Xie YF, Szeto G, Madeleine P, Tsang S. Spinal kinematics during smartphone texting - A comparison between young adults with and without chronic neck-shoulder pain. Appl Ergon. 2018;68:160-168. doi: 10.1016/j.apergo.2017.10.018.

The European Research Journal 2026;12(7):780-788 Research Article

DOI: https://doi.org/10.18621/eurj.1829791 *Chest Diseases*

Factors Affecting Mortality in COVID-19 Pneumonia Patients Treated with Tocilizumab

Selda Günaydın[1], Hayriye Bektaş Aksoy[1], İskender Aksoy[2], Abdülbaki Elmas[3], Ahmet Cumhur Dülger[4]

[1]Department of Pulmonary Medicine, Giresun University, Faculty of Medicine, Giresun, Türkiye;[2]Department of Emergency Medicine, Giresun University, Faculty of Medicine, Giresun, Türkiye;[3]Department of Internal Medicine, Aybastı State Hospital, Aybastı, Ordu, Türkiye;[4]Department of Gastroenterology, Giresun University, Faculty of Medicine, Giresun, Türkiye

Abstract:

Objective: Severe COVID-19 pneumonia is often complicated by hyperinflammation and multiorgan failure. Tocilizumab, an interleukin-6 receptor antagonist, has been used to mitigate cytokine storm; however, it's prognostic impact remains uncertain.

Methods: This retrospective cohort study included 43 adult patients with PCR-confirmed COVID-19 pneumonia who received tocilizumab at a tertiary hospital in Turkiye between April and September 2020. Patients were categorized as survivors or non-survivors. Demographic data and laboratory markers were analyzed at baseline, 24 hours, and 7 days post-treatment. Statistical analyses included Mann–Whitney U, Wilcoxon signed-rank, and chi-square tests.

Results: Non-survivors were significantly older (mean age: 72.4 vs. 64.9 years, P=0.046) and had higher levels of C-reactive protein (P=0.002), D-dimer (P=0.004), procalcitonin (P= 0.001), troponin (P=0.004), BUN (P<0.001) and creatinine (P=0.02). Survivors showed higher albumin and prognostic nutritional index (PNI) values (P=0.009 and P<0.001, respectively). On day 7, survivors exhibited increased lymphocyte, eosinophil, and platelet counts, while non-survivors had persistent neutrophilia, leukocytosis, and elevated neutrophil-to-lymphocyte ratios (P<0.05). Dynamic biomarker trends suggested ongoing inflammation and prothrombotic states among non-survivors.

Conclusion: Tocilizumab therapy may be more effective when administered early and in patients with preserved nutritional and immune function. Advanced age, renal dysfunction, elevated inflammatory markers, and poor nutritional status were significant predictors of mortality. Further prospective studies are warranted to validate these findings.

Keywords: COVID-19, Viral Pneumonia, Tocilizumab, Prognosis, Mortality

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was first identified in Wuhan, China, in December 2019. The disease rapidly spread worldwide, leading to significant morbidity and mortality due to respiratory failure. On March 11, 2020, the World Health Organization declared COVID-19 a global pandemic,

Submitted: November 25, 2025 **Accepted:** February 12, 2026 **Published Online:** February 15, 2026

How to cite this article: Günaydın S, Bektaş Aksoy H, Aksoy İ, Elmas A, Dülger AC. Factors Affecting Mortality in COVID-19 Pneumonia Patients Treated with Tocilizumab. *Eur Res J.* 2026;12(7):780-788. doi: 10.18621/eurj.1829791

Corresponding author: Selda Günaydın, MD., Assist. Prof., Phone: +90 0454 310 20 00, E-mail: seldagunaydin28@gmail.com

Available Online at https://www.eurj.org.tr

and the first case in our country was reported on the same day [1, 2].

COVID-19 exhibits a wide clinical spectrum ranging from mild symptoms to severe pneumonia and multiple organ failure. Although most patients experience a favorable course, up to 10% develop acute lung injury, acute respiratory distress syndrome (ARDS), and multiorgan involvement. Severe cases are characterized by hyaline membrane formation, inflammatory cell infiltration, and excessive cytokine release, all of which contribute to respiratory failure and mortality [3].

A hallmark of severe disease is the exaggerated immune response known as cytokine release syndrome, in which proinflammatory cytokines-particularly interleukin-6 (IL-6) play a central role [3]. Elevated IL-6 levels have been strongly associated with disease severity, respiratory failure, and mortality. Cytokine storm, considered a major driver of ARDS and multiorgan failure, significantly exacerbates the clinical course. Glucocorticoids remain the first-line therapy for macrophage activation syndrome (MAS), a condition driven by cytokine storm. In patients with inadequate response, blockade of the IL-6 receptor is recommended [1, 3].

Tocilizumab, a recombinant humanized monoclonal antibody, binds to both soluble and membrane-bound IL-6 receptors (sIL-6R and mIL-6R), thereby inhibiting IL-6–mediated signaling and reducing systemic inflammation. By targeting this pathway, tocilizumab may prevent progression of lung injury, systemic hyperinflammation, and multiorgan failure in critically ill patients [3]. The agent, already approved for the treatment of rheumatoid arthritis and cytokine release syndrome following chimeric antigen receptor T-cell therapy, has been proposed for use in COVID-19-associated hyperinflammation [1, 4].

In our country, national COVID-19 treatment guidelines, supported by current literature, recommend tocilizumab administration in patients with severe disease or in those demonstrating clinical and laboratory features of cytokine storm [5].

METHODS

This retrospective cohort study was conducted in the Department of Pulmonology at a tertiary care university hospital in Turkiye between April.15.2020 and September.30.2020, following approval by the local Ethics Committee (ODÜ KEAK 2020/226). Patients aged ≥18 years with radiologically confirmed pulmonary involvement and a positive SARS-CoV-2 polymerase chain reaction (PCR) test were eligible for inclusion. A total of 43 patients who received tocilizumab therapy were analyzed. Patients were categorized as survivors or non-survivors based on clinical outcomes. Exclusion criteria included age <18 years, refusal to provide consent for tocilizumab therapy, and unavailable medical records.

Tocilizumab was administered intravenously at a dose of 8 mg/kg (maximum 800 mg), as recommended by national guidelines, and was usually administered as a single infusion. A second dose was considered within 12–24 hours at the discretion of the treating physician if hyperinflammatory findings persisted. All patients additionally received standard care, including oxygen supplementation, dexamethasone, anticoagulation, and antiviral therapy (e.g., favipiravir), in line with institutional and national COVID-19 management protocols. All patients received a standardized corticosteroid regimen consisting of 250 mg intravenous methylprednisolone for three days administered according to the institutional protocol. Demographic data (age, sex), comorbidities, radiological findings, and treatment details were retrieved from electronic medical records. Laboratory parameters were recorded at three time points: baseline (Day 0, prior to tocilizumab administration), Day 1 (24 hours post-treatment), and Day 7 (one week post-treatment).

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 23.0 (IBM Corp., Chicago, IL, USA). The Shapiro–Wilk test was applied to assess normality of quantitative data. Comparisons of non-normally distributed variables were performed using the Mann–Whitney U test, while repeated measurements were evaluated using the Wilcoxon signed-rank test. Categorical variables were analyzed with the Pearson chi-square test. Data are presented as number (percentage), median (minimum–maximum), or mean (95% confidence interval [CI]). P-value <0.05 was considered statistically significant.

RESULTS

A total of 43 patients who received tocilizumab therapy for COVID-19 pneumonia were included in the study. The mean age of non-survivors was significantly higher than that of survivors (72.4±11.5 vs. 64.9±14.4 years, P=0.046). No significant differences were found between the two groups regarding gender distribution, concomitant favipiravir therapy, radiological findings, or length of hospital stay (Table 1).

In terms of laboratory parameters, several significant changes were observed during follow-up. On day 7 after treatment, survivors had higher lymphocyte counts (P=0.027), platelet counts (P=0.021) and eosinophil counts (P=0.020), whereas non-survivors showed marked elevations in neutrophil counts (P<0.001), white blood cell counts (P=0.004), and neutrophil-to-lymphocyte ratio (NLR) (P<0.001) (Table 2). Similarly, inflammatory and biochemical markers including C-reactive protein (CRP), procalcitonin, troponin, D-dimer, blood urea nitrogen (BUN), and creatinine levels were significantly higher in non-survivors (P<0.05). In contrast, albumin and prognostic nutritional index (PNI) levels were consistently lower among non-survivors (P<0.05) (Table 2).

Dynamic changes in biomarkers following tocilizumab administration further supported these findings. From baseline to day 7, non-survivors exhibited progressive increases in neutrophil counts, NLR, BUN, creatinine, D-dimer, procalcitonin, and sodium levels, while survivors demonstrated significant decreases in hemoglobin, hematocrit, and albumin levels (P<0.05) (Table 3). Conversely, ferritin levels decreased more prominently in survivors compared with non-survivors, suggesting a differential response to treatment.

DISCUSSION

In this study, we evaluated 43 patients with COVID-19 pneumonia who received tocilizumab therapy. Survivors and non-survivors were compared in terms of demographic, radiological, and laboratory

TABLE 1. Baseline Demographic, Clinical, and Radiological Characteristics of COVID-19 Patients Who Received Tocilizumab Therapy According to Survival Status

	Survivor	Non-survivor	P-value
Age (years)	64.9±14.4	72.4±11.5	0.046
Gender			
Male	15 (65.2%)	12 (60%)	0.724
Female	8 (34.8%)	8 (40%)	
Favipiravir			
Received	5 (21.7%)	4 (20%)	0.089
Not received	18 (78.3%)	16 (80%)	
Chest Radiology			
Ground-glass opacity	20 (87%)	18 (90%)	0.756
Fibrosis	1 (4.3%)	3 (15%)	0.230
Consolidation	4 (17.4%)	4 (20%)	0.826
Atelectasis	2 (8.7%)	4 (20%)	0.286
Pleural Effusion	3 (13%)	2 (10%)	0.756
Pulmonary Nodule	1 (4.3%)	0 (0%)	0.345
Length of stay (days)	18.5±6.3	21.2±11.6	0.779

Data are shown as mean±standard deviation or number (%) unless otherwise indicated.

TABLE 2. Comparison of Laboratory Parameters Before Treatment, on Day 1 After Treatment, and On Day 7 After Treatment in COVID-19 Patients According to Survival Status

	Before Treatment			After Treatment (1st day)			After Treatment (7th day)		
	Survivor	Non-survivor	P-value	Survivor	Non-survivor	P-value	Survivor	Non-survivor	P-value
White blood cell (×10³/µL)	10.31±6.65	7.27±4.26	0.697	11.77±13.63	11.7±5.18	0.095	10.77 ±15.99	12.08±4.55	**0.004**
Neutrophil(×10³/µL)	5.15±2.49	4.83±1.76	0.981	6.99±3.39	10.63±4.47	**0.008**	5.33±3.29	14.73±8.54	**<0.001**
Lymphocyte (×10³/µL)	1.33±1.52	1.08±0.57	0.770	1±1.19	0.75±0.39	0.551	1.25±1.43	0.68±0.3	**0.027**
Eosinophil (×10³/µL)	0.02±0.03	0.02±0.04	0.673	0.12 ± 0.28	0.03±0.06	0.138	0.18±0.16	0.09±0.09	**0.020**
Monocyte (×10³/µL)	0.73±1.04	0.31±0.16	**0.002**	0.38±0.18	0.47±0.33	0.706	0.59±0.38	0.66±0.61	0.789
Hemoglobin (g/dL)	13.5±1.4	13±2.3	0.526	12.5±1.2	12±2.3	0.232	12.6±1.4	11.2±2.5	**0.018**
Hematocrit (%)	40.5±4.3	39.4±6.4	0.408	37.7±3.9	37±6.6	0.503	38.6±3.8	35.7±7.1	0.144
Platelet (×10³/µL)	171±46	206±82	0.108	255±109	258±145	0.626	351±123	264±141	**0.021**
MPV (fL)	9.7±1	9.4±0.9	0.518	9.5±1.3	9.6±0.8	0.374	9.1±1.2	10±0.7	**0.003**
pH (unitless)	7.42±0.04	7.41±0.09	0.840	7.44±0.05	7.41±0.11	0.989	7.42±0.03	7.38±0.13	0.158
HCO_3 (mmol/L)	24.5±3.6	24.5±10.9	0.104	23.7±3	23.4±8.1	0.478	24.1±3.8	25.4±7.5	0.718
PCO_2 (mmHg)	38.7±7	35.5±10.2	0.074	36.9±6.8	35.3±7.2	0.496	38±5.9	42.8±11.4	0.211
Lactate (mmol/L)	1.8±0.6	2.2±1	0.246	1.9±0.5	2.6±1.8	0.084	2.1±0.8	3.2±4.6	0.604
Creatinine (mg/dL)	0.93±0.21	8.02±22.05	0.056	0.8±0.22	4.47±15.19	**0.009**	0.81±0.22	4.17±12.7	**0.029**
BUN (mg/dL)	37.2±14.6	57.9±31.6	**0.006**	38.1±17.5	74.4±36.9	**<0.001**	42.2±19.2	96.1±54.2	**<0.001**
AST (U/L)	33±18	55±25	**0.001**	51±44	57±48	0.428	61±44	75±86	0.874
ALT (U/L)	22±13	33±18	**0.042**	42±42	35±25	0.942	61±50	60±66	0.551
Sodium (mmol/L)	135±3	135±4	0.835	137±3	140±6	0.231	140±3	142±7	0.160
Potassium (mmol/L)	4.2±0.4	4.3±0.7	0.494	4.3±0.5	3.9±0.7	0.087	4.5±0.5	4.1±0.7	**0.025**
CRP (mg/dL)	77.86±5.89	103.26±72.16	0.075	146.75±99.89	122.64±76.91	0.495	50.47±49.2	104.94±62.1	**0.002**
D-dimer (ng/mL)	584±663	1040±1062	0.057	1917±2420	3766±2526	**0.004**	3127±2923	4392±2709	0.084
Troponin (ng/mL)	0.074±0.289	0.074±0.199	0.156	0.043±0.152	0.063±0.112	**0.006**	0.037±0.115	0.159±0.413	**0.004**
Ferritin (ng/mL)	569±497	865±718	0.144	977±572	1281±1028	0.442	823±585	1046±639	0.180
Procalcitonin (ng/mL)	0.19±0.3	0.39±0.31	**0.001**	0.19±0.31	0.40±0.32	**0.001**	0.2±0.26	0.88±1.43	**0.004**
Uric Acid (mg/dL)	4.9±1.9	5.1±1.9	0.850	4.3±1.8	4.4±2.4	0.944	4.2±1.5	4.6±3.1	0.752
Albumin (mg/L)	36.36±6.05	32.83±6.18	0.079	31.53±7.33	28.94±5.53	0.058	32.29±4.75	28.29±5.01	**0.009**
PNI	42.79±8.23	37.38±6.93	**0.042**	36.19±4.59	31.18±6.13	**0.027**	37.86±8.26	29.82±3.94	**<0.001**
NLR	8.15±10.92	6.06±4.32	0.706	9.49±7.34	16.17±10.14	**0.008**	4.14±4.52	15.17±9.56	**<0.001**

Data are shown as mean±standard deviation. NLR, neutrophil-lymphoctye ratio; CRP, C-reactive protein; MPV, mean platelet volume; PNI, prognostic nutritional index.
Statistically significant P-values are shown in bold.

TABLE 3. **Changes in Laboratory Parameters From Baseline to Day 1 and Day 7 After Tocilizumab Treatment in COVID-19 Patients (Survivor vs. Non-Survivor)**

	After Treatment (1st day)				After Treatment (7th day)			
	Survivor	**P-value**	**Non-survivor**	**P-value**	**Survivor**	**P-value**	**Non-survivor**	**P-value**
White blood cell (×10³/µL)	1.46±5.07	0.182	4.44±6	**0.004**	0.46±4.25	0.607	4.81±5.74	**0.001**
Neutrophil (×10³/µL)	0.1±0.29	0.100	0.01±0.08	0.573	0.17±0.16	**<0.001**	0.07±0.09	**0.003**
Lymphocyte (×10³/µL)	-0.32±0.47	**0.003**	-0.33±0.47	**0.005**	-0.08±0.58	0.540	-0.39±0.47	**0.001**
Eosinophil (×10³/µL)	-0.35±1	0.105	0.16±0.27	**0.020**	-0.14±0.97	0.501	0.34±0.53	**0.009**
Monocyte (×10³/µL)	1.35±11.87	0.592	10.11±8.66	**<0.001**	-4±8.97	**0.044**	9.11±8.88	**<0.001**
Hemoglobin (g/dL)	-1.01±1.04	**<0.001**	-1.02±1.36	**0.003**	-0.86±1.25	**0.003**	-1.83±1.93	**<0.001**
Hematocrit (%)	-2.85±4.12	**0.003**	-2.4±3.79	**0.011**	-1.89±4.33	**0.048**	-3.61±4.64	**0.002**
Platelet (×10³/µL)	84.48±98.92	**<0.001**	51.85±130.87	0.092	180.04±105.31	**<0.001**	58.15±129.87	0.060
MPV (fL)	-0.14±0.86	0.432	0.19±0.66	0.226	-0.57±0.88	**0.005**	0.54±0.74	**0.004**
pH (unitless)	0.01±0.05	0.247	0±0.11	1.000	0±0.05	0.962	-0.03±0.16	0.340
HCO_3 (mmol/L)	-0.44±3.01	0.539	-1.14±6.45	0.441	0.73±5.08	0.574	0.87±8.52	0.655
PCO_2 (mmHg)	-1.5±6.1	0.311	-0.23±9.85	0.920	0.14±9.27	0.951	7.28±15.57	0.050
Lactate (mmol/L)	0.05±0.37	0.542	0.44±1.79	0.287	0.3±0.84	0.173	1.03±4.72	0.339
Creatinine (mg/dL)	-0.13±0.15	**0.001**	-3.54±20.97	0.459	-0.12±0.17	**0.002**	-3.85±20.39	0.409
BUN (mg/dL)	0.87±14.93	0.783	16.5±28.83	**0.019**	5±18.25	0.202	38.2±47.01	**0.002**
AST (U/L)	18.78±46.23	0.064	1.8±45.89	0.863	28.13±35.87	**0.001**	20.05±89.8	0.331
ALT (U/L)	19.78±40.36	**0.028**	2.3±23.03	0.660	38.91±46.1	**0.001**	27.45±67.5	0.085
Sodium (mmol/L)	2.35±3.42	**0.003**	5.5±7.16	**0.003**	4.91±4.04	**<0.001**	7.6±8.31	**0.001**
Potassium (mmol/L)	0.15±0.55	0.197	-0.42±0.8	**0.031**	0.37±0.62	**0.009**	-0.29±0.91	0.171
CRP (mg/dL)	68.89±101.53	**0.004**	19.38±81.23	0.299	-27.39±79.95	0.115	1.68±86.47	0.932
D-dimer (ng/mL)	1332.87±2267.42	**0.010**	2725.95±2876.83	**<0.001**	2543.57±2648.17	**<0.001**	3352.9±3056.31	**<0.001**
Troponin (ng/mL)	-0.03±0.33	0.661	-0.01±0.2	0.803	-0.04±0.32	0.577	0.09±0.46	0.419
Ferritin (ng/mL)	408.06±464.18	**<0.001**	416.55±733.7	**0.020**	254±367.62	**0.003**	180.92±509.1	0.129
Procalcitonin (ng/mL)	1.84±4.01	**0.038**	5.8±4.09	**<0.001**	0.19±3.13	0.775	9.9±8.6	**<0.001**
Uric Acid (mg/dL)	-0.67±2.24	0.174	-0.62±2.53	0.343	-0.97±1.8	**0.023**	-0.94±2.66	0.227
Albumin (mg/L)	-4.42±8.11	**0.018**	-3.85±5.3	**0.007**	-3.83±6.03	**0.009**	-4.81±5.55	**0.003**
PNI	-5.3±5.18	**0.001**	-6.08±3.37	**<0.001**	-4.51±7.2	**0.017**	-7±6.08	**0.001**
NLR	1.35±11.87	0.592	10.11±8.66	**<0.001**	-4±8.97	**0.044**	9.11±8.88	**<0.001**

Data are shown as mean±standard deviation. NLR, neutrophil-lymphoctye ratio; CRP, C-reactive protein; MPV, mean platelet volume; PNI, prognostic nutritional index.
Statistically significant P-values are shown in bold

parameters. Our findings demonstrated that mortality was significantly associated with advanced age, renal dysfunction reflected by elevated BUN and creatinine levels, impaired poor nutritional status, indicated by low PNI and albumin levels, increased inflammatory markers (CRP, D-dimer, and procalcitonin), cardiac injury (elevated troponin), and hematological abnormalities (lymphopenia, elevated NLR, and thrombocytopenia).

Although the role of tocilizumab in COVID-19 pneumonia remains controversial, most studies highlight its potential benefit in attenuating the hyperinflammatory response. Xu *et al*. [1] reported that tocilizumab led to rapid improvements in fever, oxygenation, and radiological findings in patients with severe disease [1]. Similarly, Luo *et al*. [6] observed significant reductions in CRP and ferritin levels following treatment, though additional interventions were required in some patients. Zhang *et al*. [3] suggested that IL-6 blockade may be critical in suppressing cytokine storm and reducing mortality.

Our results are also consistent with large-scale randomized trials such as REMAP-CAP and RECOVERY, which demonstrated that tocilizumab reduced mortality, particularly when administered early in the disease course, in patients with high inflammatory burden and without advanced organ failure [7, 8]. In contrast, the randomized controlled trial by Stone *et al*. [4] did not show a significant benefit in preventing intubation or death, which may be attributed to lower baseline disease severity and delayed initiation of therapy, further emphasizing the importance of timing in the effectiveness of IL-6 receptor blockade.

The prognostic significance of nutritional status in COVID-19 has gained increasing attention. In our study, both PNI and serum albumin levels were significantly lower in non-survivors at baseline and during follow-up. Consistent with observations in the general COVID-19 pneumonia population, mortality was also higher among patients with lower PNI and albumin levels despite receiving tocilizumab therapy. These findings support the hypothesis proposed by Soetedjo *et al*. [9] who reported that hypoalbuminemia was independently associated with increased mortality, even when inflammatory markers improved. Similarly, another study evaluating nutritional biomarkers in COVID-19 emphasized the importance of including nutritional support alongside immunomodulatory therapy to optimize clinical outcomes [10].

Several retrospective studies suggest that tocilizumab may improve survival, particularly in critically ill patients. Pehlivanlar Küçük *et al*. [11] reported that tocilizumab treatment reduced all-cause mortality and improved survival in intensive care settings. Kaya *et al*. [12] further demonstrated that the survival benefit was most pronounced when tocilizumab was initiated early, within the first 6.5 days of hospitalization.

Conversely, some studies have shown no significant effect on long-term mortality. These discrepancies likely reflect differences in patient selection, comorbidity profiles, and, most importantly, the timing of tocilizumab administration [12].

The Scottish multicenter retrospective cohort study conducted by MacGregor *et al*. [13] provides valuable insights in this regard. The investigators compared early (within the first 2 days) versus late (between days 2 and 7) administration of tocilizumab and reported a 90-day mortality rate of 22% in the early group compared with 45% in the late group. Late administration increased the risk of death by more than threefold (adjusted OR: 3.33; 95% CI: 1.29–8.54; P=0.012). Consistent results were also observed for 28-day and 180-day mortality, underscoring the critical importance of treatment timing [13].

In our study, elevated CRP, D-dimer, NLR, procalcitonin, and troponin levels emerged as key predictors of mortality. These biomarkers reflect both systemic inflammation and organ dysfunction, and may help identify patients most likely to benefit from tocilizumab. Similarly, MacGregor *et al*. [13] highlighted that elevated CRP levels, in addition to treatment timing, were strongly associated with prognosis.

The marked increase in D-dimer levels among non-survivors despite tocilizumab therapy suggests the presence of an underlying prothrombotic state insufficiently controlled by IL-6 inhibition alone. Some studies have shown that, D-dimer levels remained elevated in patients receiving tocilizumab, which was explained on the basis of the lack of an effect of tocilizumab on coagulation [14, 15]. In another study evaluating D-dimer levels with tocilizumab treatment, no significant difference was

found in D-dimer levels in patients treated with tocilizumab, but changes were observed in FXIII levels. A relationship was established between these two coagulation parameters. However, D-dimer levels were found to be high in all patients, and tocilizumab treatment was reported to have no significant effect on D-dimer levels [16].

In our study, the mean age of non-survivors was significantly higher, consistent with previous evidence identifying age as one of the strongest predictors of mortality. Moreover, the higher PNI observed in survivors suggests that nutritional status and immune system reserves play an important role in determining prognosis. Similarly, elevated D-dimer, troponin, and procalcitonin levels were associated with coagulopathy, cardiac injury, and secondary infection burden, respectively, indicating that these biomarkers may be valuable in predicting mortality.

In conclusion, both our findings and existing national and international literature demonstrate that the clinical benefit of tocilizumab depends not only on its administration but also on early initiation, appropriate patient selection, and careful consideration of comorbidities.

Strengths and Limitations

This study has several notable strengths. At first, unlike many previous reports that focused only on baseline laboratory parameters, our study evaluated dynamic changes in inflammatory, hematological, and biochemical markers at three predefined time points (baseline, day 1, and day 7) following tocilizumab administration. This longitudinal biomarker assessment provides a more comprehensive understanding of treatment response rather than relying on static measurements. On the other hand, all patients received a standardized corticosteroid protocol (250 mg intravenous methylprednisolone for three days) in addition to tocilizumab, minimizing therapeutic heterogeneity and allowing a more consistent evaluation of IL-6 receptor blockade in a real-world clinical setting. And, the study assessed not only classical inflammatory markers (CRP, D-dimer, ferritin, NLR), but also included nutritional and immunological indices such as albumin and Prognostic Nutritional Index (PNI). The integration of nutritional status into the evaluation of immunomodulatory therapy has a significant role in COVID-19 research. The analysis of dynamic biomarkers between survivors and non-survivors allowed identification of early indicators of poor prognosis, such as progressive elevation of neutrophils, NLR, BUN, and procalcitonin. This provides clinically applicable insights for early risk stratification after tocilizumab therapy. Finally, the study reflects real-world clinical practice, as treatment decisions were made according to national guidelines and institutional protocols, increasing external applicability to similar tertiary care settings.

This study has several limitations. First, the absence of a non-tocilizumab control group precludes direct conclusions regarding the comparative efficacy of tocilizumab therapy; however, the primary objective of the study was to identify factors associated with mortality among patients selected for and treated with tocilizumab. Second, the retrospective and single-center design inherently limits the generalizability of our findings. Although all patients received standardized institutional treatment protocols, uncontrolled confounding factors may have influenced clinical outcomes and biomarker dynamics. Detailed respiratory severity parameters, including PaO_2/FiO_2 ratios and formal ARDS classification, were not consistently available due to the retrospective design. Third, the relatively small sample size (n=43) may have reduced statistical power to detect subtle associations, particularly in subgroup analyses. The imbalance between survivor and non-survivor groups may also have introduced bias in intergroup comparisons. Fourth, while multiple biomarkers were analyzed, cytokine levels such as IL-6 were not routinely measured after treatment due to logistical and economic constraints. This limited our ability to directly assess the immunological impact of tocilizumab.

CONCLUSION

Tocilizumab alone may not be sufficient to determine clinical outcomes in COVID-19 pneumonia; rather, its effectiveness appears to depend on appropriate patient selection, underlying comorbidities, and, most importantly, the timing of administration. Our findings suggest that tocilizumab is most beneficial when initiated early in patients with a high inflammatory

burden and preserved immune and nutritional status. Among patients treated with tocilizumab, mortality was primarily influenced by treatment timing, while nutritional status, age, and inflammatory markers should be carefully considered when selecting candidates for therapy. Future prospective, randomized controlled trials with larger cohorts are warranted to confirm these findings and to better define optimal patient selection and treatment strategies.

Ethics Approval and Consent to Participate

This study was approved by the Ordu University Clinical Research Ethics Committee. (Decision No: 2020/22-226; date: 27.10.2020). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Given the retrospective nature of the study, patient consent was not required, but data were anonymized to protect patient confidentiality.

Data Availability

All data generated or analyzed during this study are included in this published article. The data that support the findings of this study are available on request from the corresponding author, upon reasonable request.

Authors' Contribution

Study Conception: SG, HBA, İA; Study Design: SG, HBA, İA; Supervision: SG; Funding: N/A; Materials: AE, ACD, SG; Data Collection and/or Processing: AE, ACD, SG; Statistical Analysis and/or Data Interpretation: İA, HBA; Literature Review: SG, HBA; Manuscript Preparation: SG; and Critical Review: SG.

Conflict of Interest

The author(s) disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The author(s) disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgments

The authors would like to express their sincere gratitude to the medical and nursing staff of Giresun University Training and Research Hospital for their dedicated care and efforts in managing COVID-19 patients during the pandemic. We also thank the data management unit and biostatistics support staff for their valuable assistance during data collection and analysis.

Generative Artificial Intelligence Statement

The author(s) declare that no artificial intelligence-based tools or applications were used during the preparation process of this manuscript. The all content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles. AI assistance was only used during grammar correction after the article had been written.

Editor's Note

All statements made in this article are solely those of the authors and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Xu X, Han M, Li T, et al. Effective treatment of severe COVID-19 patients with tocilizumab. Proc Natl Acad Sci U S A. 2020;117(20):10970-10975. doi: 10.1073/pnas.2005615117.
2. Zhonghua Liu Xing Bing Xue Za Zhi. Epidemiology Working Group for NCIP Epidemic Response, Chinese Center for Disease Control and Prevention. [The epidemiological characteristics of an outbreak of 2019 novel coronavirus diseases (COVID-19) in China]. 2020 Feb 10;41(2):145-151. doi: 10.3760/cma.j.issn.0254-6450.2020.02.003. [Article in Chinese]
3. Zhang C, Wu Z, Li JW, Zhao H, Wang GQ. Cytokine release syndrome in severe COVID-19: interleukin-6 receptor antagonist tocilizumab may be the key to reduce mortality. Int J Antimicrob Agents. 2020;55(5):105954. doi: 10.1016/j.ijantimicag.2020.105954.
4. Stone JH, Tuckwell K, Dimonaco S, et al. Trial of Tocilizumab in Giant-Cell Arteritis. N Engl J Med. 2017;377(4):317-328. doi: 10.1056/NEJMoa1613849.
5. Republic of Turkey Ministry of Health. Anticytokine-antiinflammatory treatments, coagulopathy management. Ankara: Ministry of Health Publications; 2021.
6. Luo P, Liu Y, Qiu L, Liu X, Liu D, Li J. Tocilizumab treatment in COVID-19: A single center experience. J Med Virol. 2020;92(7):814-818. doi: 10.1002/jmv.25801.
7. REMAP-CAP Investigators; Gordon AC, Mouncey PR, Al-Beidh F, et al. Interleukin-6 Receptor Antagonists in Critically

Ill Patients with Covid-19. N Engl J Med. 2021;384(16):1491-1502. doi: 10.1056/NEJMoa2100433.
8. RECOVERY Collaborative Group. Tocilizumab in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial. Lancet. 2021;397(10285):1637-1645. doi: 10.1016/S0140-6736(21)00676-0.
9. Soetedjo NNM, Iryaningrum MR, Damara FA, et al. Prognostic properties of hypoalbuminemia in COVID-19 patients: A systematic review and diagnostic meta-analysis. Clin Nutr ESPEN. 2021;45:120-126. doi: 10.1016/j.clnesp.2021.07.003.
10. Anzo FM, Buan-Mayo M. Nutritional biomarkers as predictors of clinical outcomes between COVID-19 severity groups in a tertiary government hospital. Clin Nutr ESPEN. 2023;53:134-143. doi: 10.1016/j.clnesp.2022.12.005.
11. Pehlivanlar Küçük M, Küçük AO, Pehlivanlar A, et al. Effect of tocilizumab on intensive care patients with Covid-19 pneumonia, a retrospective cohort study. Turk J Med Sci. 2022;52(1):39-49. doi: 10.3906/sag-2106-42.
12. Kaya H, Öksüzler Kızılbay G, Ilgazlı AH, Özgür EG. Tocilizumab in hospitalized patients with severe COVID-19 pneumonia: A single-center observational study. Eurasian J Pulmonol 2024;26(1):41-50. doi: 10.14744/ejp.2023.4009.
13. MacGregor F, Oprey A, Caulfield C, MacTavish P, Lowrie R, Henderson P. Does timing of tocilizumab administration affect mortality in COVID-19? A Scottish multicentre retrospective cohort study. BMJ Open Respir Res. 2024;11(1):e002264. doi: 10.1136/bmjresp-2023-002264.
14. Toniati P, Piva S, Cattalini M, et al. Tocilizumab for the treatment of severe COVID-19 pneumonia with hyperinflammatory syndrome and acute respiratory failure: A single center study of 100 patients in Brescia, Italy. Autoimmun Rev. 2020;19(7):102568. doi: 10.1016/j.autrev.2020.102568.
15. Tosun M, Ölmez H. Effect of Tocilizumab Use on Mortality in COVID-19 Patients Admitted to Intensive Care Unit. Duzce Med J. 2022;24(3):227–234. doi: 10.18678/dtfd.1108303.
16. Cibis C, Adam EH, Gatzke F, Rauchfuss S, Roth S, Miesbach W. Coagulation Biomarkers and Mortality in Severe COVID-19: The Prognostic Value of Factor XIII Decline and Elevated D-Dimers. Clin Appl Thromb Hemost. 2025;31:10760296251364266. doi: 10.1177/10760296251364266.

The European Research Journal 2026;12(7):789-797 Research Article

DOI: https://doi.org/10.18621/eurj.1840243 *Orthopedics and Traumatology*

Intra-Articular Platelet-Rich Plasma, Hyaluronic Acid, and Mesenchymal Stem Cell for Knee Osteoarthritis (2000-2025): A Multi-Database Bibliometric Analysis

Ruhat Ünlü[1], Hasan Emirhan Usta[2]

[1]Department of Orthopedics and Traumatology, İstanbul Atlas University, Faculty of Medicine, İstanbul, Türkiye; [2]Department of Orthopedics and Traumatology, Gebze Fatih State Hospital, Kocaeli, Türkiye

Abstract:

Objective: Biologic intra-articular injections, including platelet-rich plasma (PRP), hyaluronic acid (HA), and mesenchymal stem cell (MSC) based therapies, are increasingly investigated for knee osteoarthritis (KOA). However, no comprehensive, multidatabase bibliometric study has compared publication dynamics, thematic evolution, and international collaboration patterns across these modalities.

Methods: We searched PubMed, Web of Science, and Scopus for studies published between 2000 and 2025 using standardized terms for KOA and intra-articular PRP/HA/MSC. Records were merged and deduplicated in R (bibliometrix). Descriptive bibliometrics (publication counts, country output) and science mapping techniques (co-authorship networks, thematic co-word analysis) were applied. Figures included annual publication trends, geographic distribution, collaboration networks, thematic networks, and a PRISMA flow diagram.

Results: After duplication, 2,291 unique publications were included. PRP-related research grew steeply after 2012 and remained dominant; MSC-related publications expanded rapidly after 2015; HA output was modest and relatively stable. A total of 82 countries contributed, led by the United States (n=2,074) and China (n=1,559), followed by Italy, France, and Spain. Strong bilateral collaborations were observed between Italy and the United Kingdom, the USA–Canada, and Italy–Switzerland. Thematic analysis identified hydrogel, intra-articular HA, and rehabilitation as motor themes, while extracellular vesicles and cartilage regeneration emerged as hotspots.

Conclusion: Research on biologic intra-articular injections for KOA has grown substantially, with PRP and MSC driving recent expansion. The field is geographically broad but dominated by North America, China, and Western Europe. Hotspot clusters suggest a shift toward cell-free biologics and regenerative strategies. Our open, reproducible workflow provides a transparent basis for ongoing bibliometric monitoring and evidence synthesis.

Keywords: Knee Osteoarthritis, Platelet-Rich Plasma, Hyaluronic Acid, Mesenchymal Stem Cells, Intra-Articular Injection, Bibliometric Analysis, Regenerative Therapy

Submitted: December 11, 2025 **Accepted:** January 17, 2026 **Published Online:** January 20, 2026

How to cite this article: Ünlü R, Usta HE. Intra-Articular Platelet-Rich Plasma, Hyaluronic Acid, and Mesenchymal Stem Cell for Knee Osteoarthritis (2000-2025): A Multi-Database Bibliometric Analysis. *Eur Res J.* 2026;12(7):789-797. doi: 10.18621/eurj.1840243

Corresponding author: Ruhat Ünlü, MD., Asisst. Prof., Phone: +90 212 440 40 00, E-mail: ruhat_unlu@hotmail.com, ruhat.unlu@atlas.edu.tr

Available Online at https://www.eurj.org.tr

Knee osteoarthritis (KOA) is one of the most common causes of pain and disability in adults worldwide. Its prevalence continues to increase with population aging and obesity [1, 2]. KOA is now recognized as a whole-joint disorder involving cartilage, subchondral bone, synovium, and periarticular soft tissues [1]. Standard management is mainly conservative, focusing on symptom relief through exercise, physical therapy, weight management, and intra-articular injections [2].

Among injectable options, hyaluronic acid (HA) has been widely used for decades as a viscosupplementation therapy [3]. However, its efficacy has been debated because of heterogeneity in molecular formulations and trial results [3]. Platelet-rich plasma (PRP) has gained attention during the last decade. Several randomized trials and meta-analyses suggest that PRP may provide superior pain relief and functional improvement compared with HA [4, 5]. Still, results vary due to differences in preparation methods, dosing regimens, and concomitant treatments [6]. Mesenchymal stem cell (MSC)–based approaches, including exosome-derived therapies, represent an emerging area with promising anti-inflammatory and chondroprotective effects [7–9].

Despite the growing body of clinical and preclinical evidence, few studies have comprehensively mapped the scientific landscape of biologic injections for KOA. Some bibliometric analyses have examined PRP alone [8] or intra-articular therapies in general [9], but no study has compared PRP, HA, and MSC simultaneously across multiple databases and a long-time frame. This gap limits our understanding of publication dynamics, thematic hotspots, and international collaboration networks.

This study aimed to merge PubMed, Web of Science, and Scopus records from 2000 to 2025 and provide a comprehensive bibliometric analysis of intra-articular PRP, HA, and MSC for KOA. We sought to describe temporal trends, identify research themes, and map geographic collaboration patterns to guide future research directions.

METHODS

Search Strategy and Data Sources

We conducted a comprehensive literature search in PubMed, Web of Science (WoS Core Collection), and Scopus from January 1, 2000, to September 9, 2025. The search strategy combined terms for KOA with PRP, HA, MSCs, and intra-articular injection. Search syntax was adapted for each database. Records were exported in native formats (.nbib for PubMed, .txt for WoS, csv for Scopus). The bibliometric analysis was conducted using the Bibliometrix R package and

Its web interface, Biblioshiny [10].

Eligibility Criteria

Records were included if they reported on intra-articular PRP, HA, or MSC therapies for KOA. We applied no restriction on study design or article type. Non-English articles, duplicates across databases, and unrelated topics were excluded. No human or animal ethics approval was required, as only published literature was analyzed.

Data cleaning and integration

All records were merged in RStudio (R version 2025.05.1). Initial totals were 1,858 (PubMed), 3,012 (WoS), and 2,913 (Scopus). After deduplication, 2,291 unique publications remained. Deduplication involved DOI and title-based screening, supplemented by manual checks for residual duplicates. Records were harmonized to standard fields, including title, abstract, keywords, author information, country, and year of publication.

Bibliometric Analysis

We performed descriptive analyses (annual output, top countries, prolific collaborations) and science mapping. Country-level output and co-authorship networks were constructed. International collaboration networks were visualized using Louvain clustering. Thematic evolution was explored by co-word analysis based on Author Keywords (DE) and Keywords Plus (ID). Results were displayed as temporal trends, world maps, collaboration networks, thematic maps, and co-word clusters.

Reporting Framework

The review process followed the PRISMA 2020 guidelines for transparent reporting of literature searches [11]. A PRISMA-style flow diagram was

included to depict identification, screening, and inclusion of records.

Statistical Analysis

Statistical analysis in this study was conducted within the framework of bibliometric and network-based methods rather than inferential clinical statistics. All analyses were performed using the bibliometrix package (version 4.5.1) in R (R Foundation for Statistical Computing, Vienna, Austria) and its web-based interface, Biblioshiny. Descriptive bibliometric indicators were calculated to summarize publication characteristics, including annual publication counts, country-level productivity, and collaboration frequencies. Publication trends over time were analyzed using absolute frequencies and temporal distributions without hypothesis testing, as recommended for bibliometric studies. Network analyses were applied to evaluate scientific collaboration and thematic structure. Country-level co-authorship networks were constructed based on author affiliation data. Network visualization and clustering were performed using the Louvain community detection algorithm to identify major collaboration clusters. Network nodes represent countries, node size reflects publication volume, and edge thickness indicates collaboration strength. Thematic analysis was conducted using co-word analysis based on Author Keywords (DE) and Keywords Plus (ID). Keyword co-occurrence matrices were generated, and thematic maps were produced using Callon's centrality and density measures to classify themes into motor, basic, emerging, and declining categories. Thematic evolution was assessed descriptively across the entire study period. No inferential statistical tests, effect size calculations, or p-values were applied, as the aim of the study was to map research trends, thematic development, and collaboration patterns rather than to test causal hypotheses. Results are presented descriptively through tables, figures, and network visualizations.

RESULTS

Search Results

A total of 1,858 records were identified from PubMed, 3,012 from Web of Science, and 2,913 from Scopus. After merging and deduplication, 2,291 unique records were included in the final dataset (Figure 1, PRISMA flow diagram).

Flow of records through identification, deduplication, screening, and inclusion. A total of 1,858 (PubMed), 3,012 (Web of Science), and 2,913 (Scopus) records were identified; after deduplication, 2,291 unique records were included in the bibliometric analysis.

Annual Trends

Publication output increased steadily after 2010. PRP-related studies experienced a steep increase after 2012 and remained dominant. MSC-related

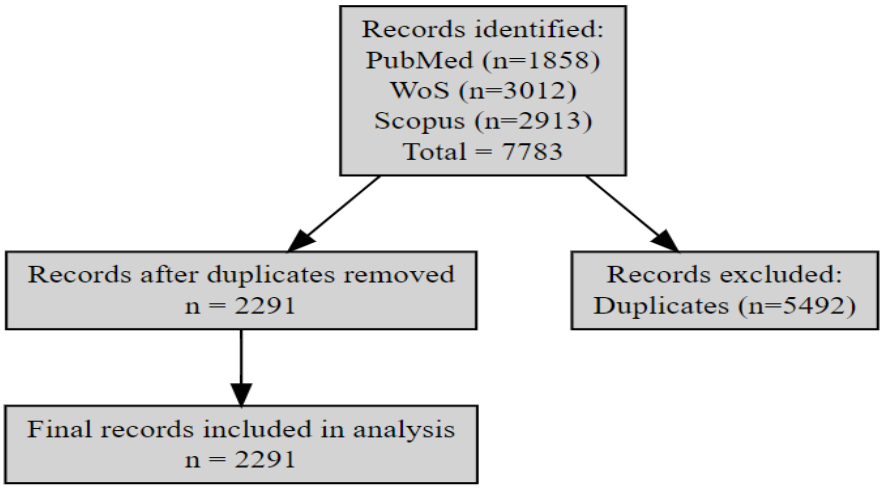

FIGURE 1. PRISMA flow diagram.

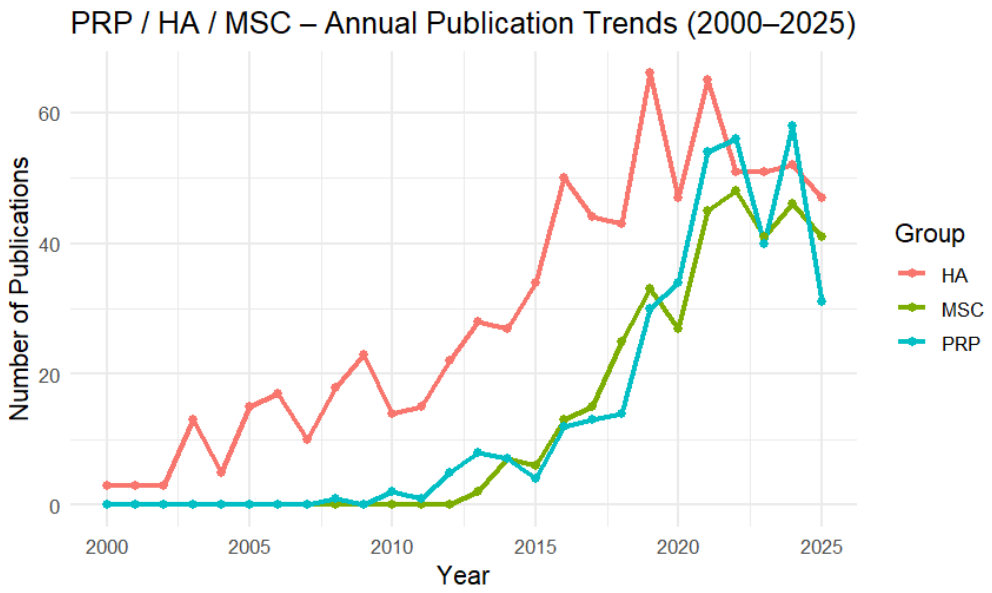

FIGURE 2. Platelet-rich plasma (PRP) vs Hyaluronic acid (HA) vs Mesenchymal stem cell (MSC) – Annual publication trends (2000–2025).

publications rose sharply after 2015. HA-related output was modest and relatively stable throughout the period (Figure 2). As shown in Figure 2, PRP publications increased markedly after 2012, whereas HA output remained relatively stable.

Geographic Distribution

In total, eighty-two countries contributed to the field. Among them, the United States accounted for the highest number of publications (n=2,074), followed by China (n=1,559), Italy (n=846), France (n=493), and Spain (n=379) (Table 1). Other productive contributors included Korea, the United Kingdom, Japan, Canada, and Australia (Figure 3).

International Collaborations

Collaboration networks revealed strong cross-border ties. The most frequent co-authorship links were Italy–United Kingdom (n=44), USA–Canada (n=41), Italy–Switzerland (n=40), Italy–Germany (n=35), and USA–United Kingdom (n=33) (Table 2). These partnerships reflect the central role of Western Europe and North America in shaping the field (Figure 4).

Country-level co-authorship network. Node size corresponds to publication volume; edge thickness reflects collaboration strength. Labels highlight the most productive countries; strong collaborations were observed between Italy–United Kingdom, USA–Canada, and Italy–Switzerland.

Thematic Structure

Thematic co-word mapping identified several motor themes, including intra-articular hyaluronic

TABLE 1. Top 10 Countries by Publication Volume (2000–2025)

Rank	Country	Publications (n)
1	USA	2,074
2	China	1,559
3	Italy	846
4	France	493
5	Spain	379
6	Korea	339
7	United Kingdom	298
8	Japan	295
9	Canada	271
10	Australia	247

Top 10 countries ranked by the number of publications included in the final bibliometric dataset between 2000 and 2025. "Publications (n)" indicates the count of included documents per country based on author affiliation–country assignment. Note: Counts refer to the deduplicated dataset.

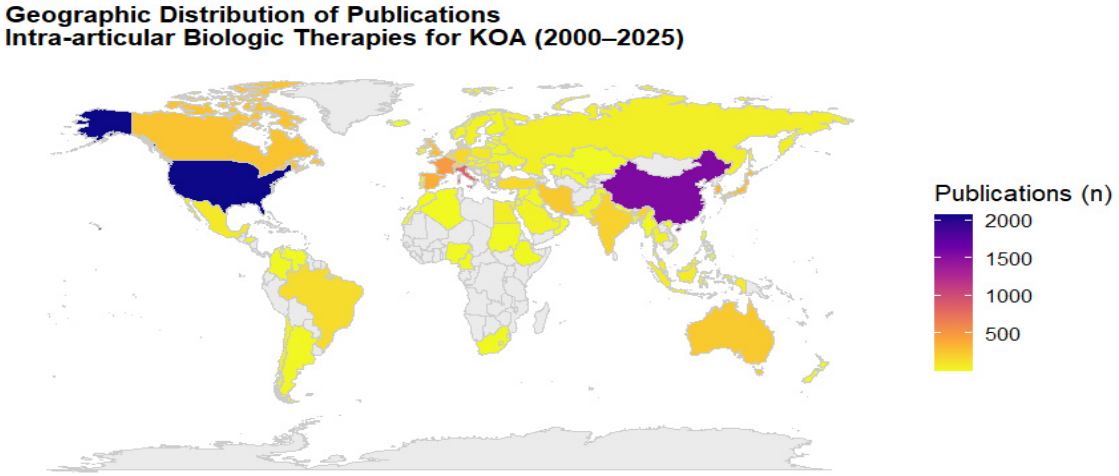

FIGURE 3. Geographic Distribution of Publications (2000–2025). Choropleth map of country-level publication counts. Darker shading indicates higher output. The United States and China contributed the largest volumes, followed by Italy, France, and Spain.

acid, rehabilitation, and hydrogel research. Emerging clusters were observed around extracellular vesicles and cartilage regeneration, suggesting new research directions in regenerative therapy (Figure 5).

Co-occurrence network derived from Keywords Plus (ID). Communities detected using the Louvain algorithm. Node size indicates term degree (centrality), and edge transparency reflects co-occurrence weight. Labels show the most central terms within clusters.

TABLE 2. Top 10 International Collaborations (2000–2025)

Rank	Country pair	Co-authored publications (n)
1	Italy – United Kingdom	44
2	USA – Canada	41
3	Italy – Switzerland	40
4	Italy – Germany	35
5	USA – United Kingdom	33
6	USA – Italy	32
7	Italy – Spain	32
8	France – Belgium	32
9	Italy – France	28
10	USA – China	25

Country pairs with the highest number of co-authored publications in the final bibliometric dataset between 2000 and 2025. "Co-authored publications (n)" counts documents listing at least one author from each country in the pair.

Note: Collaboration strength is reflected by the number of joint publications.

DISCUSSION

This multidatabase bibliometric study shows a consistent rise in publications on intra-articular biologic injections for knee osteoarthritis. Output accelerated after 2012, with PRP accounting for the largest share, followed by MSC and HA. These bibliometric patterns mirror clinical findings. Meta-analyses have reported superior pain and functional outcomes for PRP compared with HA [4, 5]. Meanwhile, MSC therapies have gained interest due to their regenerative and immunomodulatory potential, as emphasized in recent systematic reviews [7–9].

The United States and China produced the highest number of publications, followed by Italy, France, and Spain. This distribution aligns with their recognized leadership in orthopedic and regenerative research [1, 2]. Italy emerged as a European hub, forming strong bilateral collaborations with the United Kingdom, Germany, Switzerland, and Spain. Such partnerships enhance visibility and accelerate clinical translation. Earlier bibliometric studies limited to PRP or intra-articular therapy did not offer this level of comparative geographic and collaborative mapping [8, 9]. Our study addresses this gap by integrating PubMed, Web

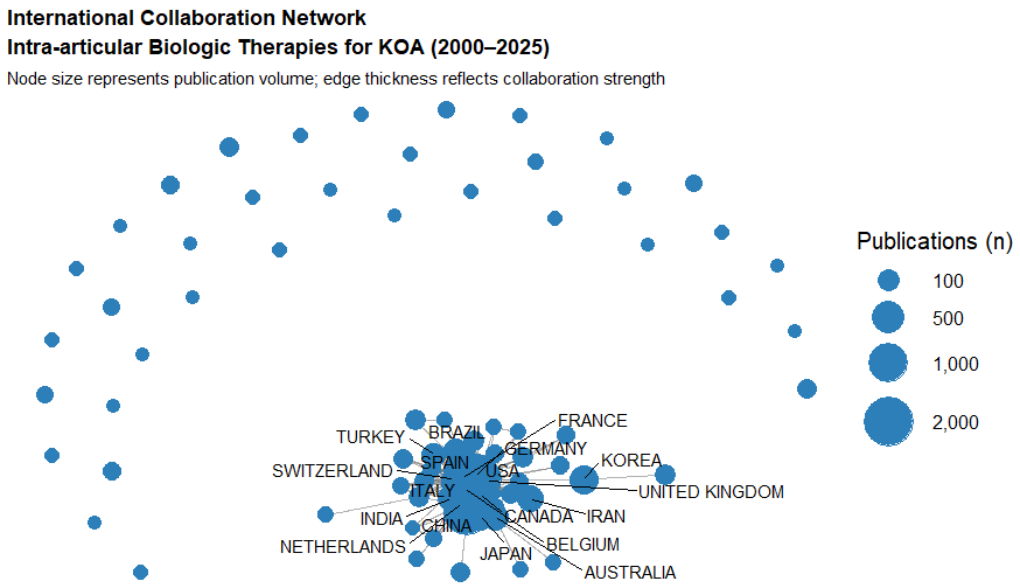

FIGURE 4. International Collaboration Network (2000–2025). Communities were identified using Louvain clustering; larger nodes represent higher publication output.).

of Science, and Scopus into a unified analysis.

Thematic signals converged on cell-free biologics and regenerative strategies. Reviews and year-in-review reports emphasize that exercise-centered rehabilitation remains the core of KOA care, and that adjunct intra-articular options should be integrated within multimodal programs rather than used in isolation [12, 13]. Within biologics, extracellular vesicles (EVs)/exosomes are emerging as prominent hotspots. Recent reviews highlight their immunomodulatory and chondroprotective actions and the translational appeal of cell-free delivery

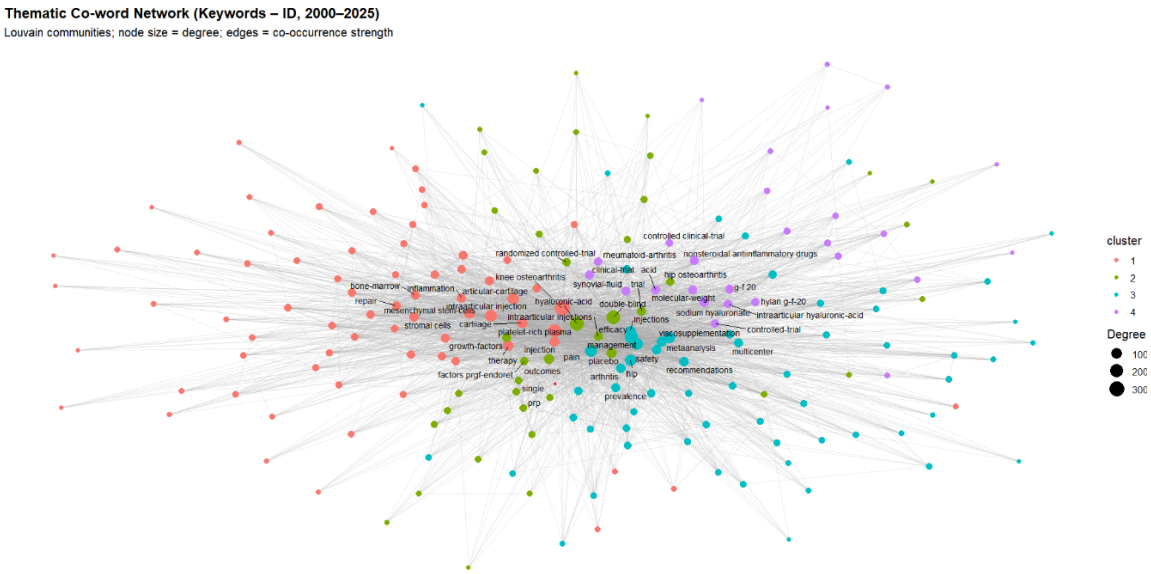

FIGURE 5. Thematic Co-word Network of Keywords (2000–2025). Node size reflects keyword centrality, while edge thickness represents co-occurrence strength between terms.

compared with whole-cell MSC therapies [14]. Preclinical synthesis shows consistent structural and pain-related benefits of MSC-derived exosomes in KOA models, supporting continued clinical development [15]. On the viscosupplementation/hydrogel axis, product-level differences in hyaluronic acid remain relevant, with evidence suggesting short-term pain reduction and potential formulation-specific effects - findings that fit our "motor theme" observation around intra-articular HA [6]. Finally, the regenerative stream points toward cartilage repair trajectories; contemporary overviews underscore both the progress and the methodological gaps that next-generation biologics must address to achieve durable structure-modifying outcomes [16]. Engineering approaches for long-acting EV platforms and targeted delivery are proposed as near-term solutions to improve persistence and joint-tissue uptake, aligning with the keyword network we observed [17].

A strength of this study is the integration of three major databases (PubMed, Web of Science, Scopus), reducing the bias of single-source coverage. The workflow followed the PRISMA framework for transparent reporting [11] and applied an open-source R package (bibliometrix) that allows reproducibility [10]. Compared with earlier bibliometric efforts that focused only on PRP [8] or general intra-articular injections [9], our approach provides a wider perspective and a standardized data-cleaning pipeline. The use of multiple visualizations annual trends, world maps, collaboration networks, and thematic mapping, offers a multidimensional view of the field. Such integration of quantitative metrics with science mapping is increasingly recognized as a best practice in bibliometric studies [18, 19].

The expansion of research on intra-articular biologics for KOA reflects the clinical need for alternatives beyond conventional analgesics and surgery. PRP and MSC dominate recent publication growth, suggesting that regenerative and cell-free approaches may increasingly shape treatment pathways. Current OARSI guidelines highlight exercise and weight management as cornerstones of care, while allowing intra-articular injections as adjunctive options in selected patients [2]. At the same time, expert consensus statements underline that biologic injections should not replace established rehabilitation programs but may complement them when standard care is insufficient [2]. Our bibliometric results show rising attention to extracellular vesicles and cartilage regeneration, indicating a translational shift that requires robust randomized trials and standardized protocols before routine clinical use [14, 15, 20]. These observations reinforce the need for closer integration between basic science, clinical research, and guideline development, and they provide the rationale for our concluding statements. From a translational perspective, the bibliometric trends identified in this analysis may help inform future research priorities rather than direct clinical decision-making. The growing emphasis on PRP standardization, MSC-derived extracellular vesicles, and hydrogel-based delivery systems highlights areas where well-designed randomized controlled trials are particularly needed. Mapping these trends may assist researchers and guideline developers in identifying gaps in evidence, harmonizing outcome measures, and designing comparative trials that address current heterogeneity in biologic injection protocols.

Strengths and Limitations

This study has several notable strengths. First, it integrates three major bibliographic databases (PubMed, Web of Science, and Scopus), which enhances coverage and reduces the risk of single-database bias compared with previous bibliometric studies focused on a single source. Second, the long observation period (2000–2025) allows for a comprehensive evaluation of temporal publication trends and thematic evolution in intra-articular biologic therapies for knee osteoarthritis. Third, the use of standardized, reproducible methods based on the bibliometrix R package enables transparent data processing, network construction, and thematic mapping, facilitating reproducibility and future updates of the analysis. Finally, combining descriptive bibliometrics with collaboration and co-word network analyses provides a multidimensional overview of research productivity, international cooperation, and emerging scientific themes.

Despite these strengths, several limitations should be acknowledged. First, database coverage bias cannot be fully excluded, as PubMed, Web of Science, and Scopus differ in journal indexing policies, regional representation, and citation practices. Although

integrating multiple databases reduces single-source bias, some relevant publications may still be underrepresented or inconsistently indexed. Second, citation lag may affect the visibility of recently published studies, particularly those from the final years of the study period, potentially underestimating their scientific impact. Third, bibliometric indicators primarily reflect publication volume, citation patterns, and network structures rather than study quality, methodological rigor, or clinical effectiveness. Therefore, the observed trends should be interpreted as indicators of research activity and thematic development, not as evidence of therapeutic superiority or clinical efficacy.

CONCLUSION

This bibliometric analysis provides a comprehensive overview of global research activity on intra-articular biologic therapies for knee osteoarthritis over the past 25 years. By integrating three major databases and systematically comparing PRP, HA, and MSC-related publications, the study delineates distinct temporal trajectories, evolving thematic structures, and international collaboration patterns. PRP has emerged as the most extensively studied modality, HA has maintained a stable research presence, and MSC-related research has expanded rapidly in recent years. Importantly, these findings reflect research dynamics and thematic evolution rather than clinical effectiveness. The identification of emerging topics such as extracellular vesicles and advanced delivery systems highlights shifting research interests toward regenerative concepts. This bibliometric framework offers a transparent basis for future evidence synthesis and ongoing monitoring of research trends in this evolving field.

Ethics Approval and Consent to Participate

This study did not require approval from an Institutional Review Board/Ethics Committee because it does not involve human participants, human-derived materials, identifiable personal data, or any interventions affecting patient care. The research is based entirely on publicly available sources and does not include any clinical data, patient records, or experimental procedures. This study analyzed data exclusively from published literature indexed in PubMed, Web of Science, and Scopus. Therefore, ethical approval and informed consent were not required.

Data Availability

All data generated or analyzed during this study are included in this published article. The data that support the findings of this study are available on request from the corresponding author, upon reasonable request.

Authors' Contribution

Study Conception: RÜ; Study Design: RÜ; Supervision: RÜ, HEU; Funding: RÜ, HEU; Materials: RÜ, HEU; Data Collection and/or Processing: RÜ, HEU; Statistical Analysis and/or Data Interpretation: RÜ, HEU; Literature Review: RÜ, HEU; Manuscript Preparation: RÜ, HEU; and Critical Review: RÜ, HEU.

Conflict of Interest

The author(s) disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The author(s) disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgment

The authors would like to acknowledge the contribution of the independent reviewers who provided feedback during the preparation of this manuscript.

Generative Artificial Intelligence Statement

The authors utilized the Bibliometrix R package for bibliometric analysis. In addition, artificial intelligence tools (ChatGPT by OpenAI and Gemini by Google) were employed exclusively for language polishing and formatting support. All scientific content, data analysis, and interpretations were conducted solely by the authors. The all content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

Editor's Note

All statements made in this article are solely those of the authors and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. Lancet. 2019;393(10182):1745-1759. doi: 10.1016/S0140-6736(19)30417-9.
2. Bannuru RR, Osani MC, Vaysbrot EE, et al. OARSI guidelines for the non-surgical management of knee, hip, and polyarticular osteoarthritis. Osteoarthritis Cartilage. 2019;27(11):1578-1589. doi: 10.1016/j.joca.2019.06.011.
3. Xing D, Wang B, Liu Q, et al. Intra-articular Hyaluronic Acid in Treating Knee Osteoarthritis: a PRISMA-Compliant Systematic Review of Overlapping Meta-analysis. Sci Rep. 2016;6:32790. doi: 10.1038/srep32790.
4. Belk JW, Kraeutler MJ, Houck DA, Goodrich JA, Dragoo JL, McCarty EC. Platelet-Rich Plasma Versus Hyaluronic Acid for Knee Osteoarthritis: A Systematic Review and Meta-analysis of Randomized Controlled Trials. Am J Sports Med. 2021;49(1):249-260. doi: 10.1177/0363546520909397.
5. Tang JZ, Nie MJ, Zhao JZ, Zhang GC, Zhang Q, Wang B. Platelet-rich plasma versus hyaluronic acid in the treatment of knee osteoarthritis: a meta-analysis. J Orthop Surg Res. 2020;15(1):403. doi: 10.1186/s13018-020-01919-9.
6. Ferkel E, Manjoo A, Martins D, Bhandari M, Sethi P, Nicholls M. Intra-articular Hyaluronic Acid Treatments for Knee Osteoarthritis: A Systematic Review of Product Properties. Cartilage. 2023;14(4):424-432. doi: 10.1177/19476035231154530.
7. Liu J, Gao J, Niu Q, Wu F, Wu Z, Zhang L. Bibliometric and visualization analysis of mesenchymal stem cells and rheumatoid arthritis (from 2012 to 2021). Front Immunol. 2022;13:1001598. doi: 10.3389/fimmu.2022.1001598.
8. Xiao Z, Chen W, Wei Z, Zhang Q, Tang G. Global trends and hotspots in the application of platelet-rich plasma in knee osteoarthritis: A bibliometric analysis from 2008 to 2022. Medicine (Baltimore). 2023;102(47):e35854. doi: 10.1097/MD.0000000000035854.
9. Lu Z, Xie L, Liu W, et al. A bibliometric analysis of intra-articular injection therapy for knee osteoarthritis from 2012 to 2022. Medicine (Baltimore). 2023;102(46):e36105. doi: 10.1097/MD.0000000000036105.
10. Aria M, Cuccurullo C. bibliometrix: An R-tool for comprehensive science mapping analysis. J Informetr. 2017;11(4):959-975. doi: 10.1016/j.joi.2017.08.007.
11. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ. 2021;372:n71. doi: 10.1136/bmj.n71.
12. Macri EM, Selles RW, Stefanik JJ, Reijman M. OARSI year in review 2023: Rehabilitation and outcomes. Osteoarthritis Cartilage. 2023;31(12):1534-1547. doi: 10.1016/j.joca.2023.08.011.
13. Lawford BJ, Bennell KL, Haber T, et al. Osteoarthritis Year In Review 2024: Rehabilitation and outcomes. Osteoarthritis Cartilage. 2024;32(11):1405-1412. doi: 10.1016/j.joca.2024.08.001.
14. Luo D, Zhu H, Li S, Wang Z, Xiao J. Mesenchymal stem cell-derived exosomes as a promising cell-free therapy for knee osteoarthritis. Front Bioeng Biotechnol. 2024;12:1309946. doi: 10.3389/fbioe.2024.1309946.
15. Wang Z, Hu Z, Niu L, Xu Y, Qi Y. Mesenchymal stem cell-derived exosomes for the treatment of knee osteoarthritis: a systematic review and meta-analysis based on rat model. Front Pharmacol. 2025;16:1588841. doi: 10.3389/fphar.2025.1588841.
16. Brittberg M. Treatment of knee cartilage lesions in 2024: From hyaluronic acid to regenerative medicine. J Exp Orthop. 2024;11(2):e12016. doi: 10.1002/jeo2.12016.
17. Drohat P, Baron M, Kaplan LD, Best TM, Kouroupis D. Long-Acting Extracellular Vesicle-Based Biologics in Osteoarthritis Immunotherapy. Bioengineering (Basel). 2025;12(5):525. doi: 10.3390/bioengineering12050525.
18. Donthu N, Kumar S, Mukherjee D, Pandey N, Lim WM. How to conduct a bibliometric analysis: An overview and guidelines. J Bus Res. 2021;133:285-296. doi: 10.1016/j.jbusres.2021.04.070.
19. Chen C. Science mapping: a systematic review of the literature. J Data Inf Sci. 2017;2(2):1-40. doi: 10.1515/jdis-2017-0006.
20. Sun C, Teng F, Xia Y. Extracellular vesicles in osteoarthritis: mechanisms, therapeutic potential, and diagnostic applications. Front Immunol. 2025;16:1595095. doi: 10.3389/fimmu.2025.1595095.

The European Research Journal 2026;12(7):798-802 Case Report

DOI: https://doi.org/10.18621/eurj.1842449 *General Surgery*

Not All That Masses Are Cancer: Pilomatrixoma Presenting as Breast Cancer

Murat Özkara[1], Şebnem Çimen[2], Mehmet Mert Hıdıroğlu[1]

[1]Department of General Surgery, University of Health Sciences, Gülhane Training and Research Hospital, Ankara, Türkiye; [2]Department of General Surgery, Akyurt State Hospital, Ankara, Türkiye

Abstract:

Pilomatrixoma is a rare benign skin neoplasm originating from the hair follicle matrix and is exceptionally uncommon in breast tissue, particularly in male patients. Due to overlapping clinical and ultrasonographic features, it is frequently misdiagnosed as breast cancer. This study aims to highlight pilomatrixoma as an important differential diagnosis of breast masses and to emphasize the role of histopathological confirmation. We report two male patients with breast lesions. A 40-year-old male presented with an infected breast lesion, and a 70-year-old male presented with a palpable breast mass. Both patients underwent clinical evaluation, ultrasonography, and tru-cut biopsy. Following histopathological confirmation of pilomatrixoma, elective surgical excision was performed. Tru-cut biopsy findings in both cases were consistent with pilomatrixoma. Ultrasonographic evaluation suggested suspicious features, leading to BI-RADS 4–5 categorization and initial concern for malignancy. Complete surgical excision with clear margins was achieved in both patients, and no complications or recurrences were observed. Pilomatrixoma is a rare benign tumor of the skin that can closely mimic breast cancer, especially on imaging. Despite its typical occurrence in the head, neck, and upper extremities and predominance in young females, it should be considered in the differential diagnosis of male breast masses. Histopathological confirmation via tru-cut biopsy is essential for accurate diagnosis, and surgical excision with clear margins is curative.

Keywords: Pilomatrixoma, Breast Mass, Breast Cancer, Benign Skin Tumor

Pilomatrixoma is a rare skin neoplasm arising from the hair follicle matrix. It mostly occurs in the head, neck, and upper extremities, and rarely on the trunk [1]. Based on a literature search conducted in the PubMed database, 22 cases involving breast tissue have been reported to date [1–20]. It generally occurs in the first or second decades of life. Pilomatrixoma – known also as Malherbe's calcified epithelioma – is often confused with breast cancer on physical examination when arising in the breast tissue [17]. The lesion can be confused with a malignancy based on radiological findings and be reported as BI-RADS 4 or 5 (Breast Imaging Reporting and Data System) [17], and presents as an asymptomatic, mobile, and slow-growing nodule [21]. We describe here two cases of this rare disease that presented in an unusual location within the breast tissue. This case report has been compiled in accordance with the SCARE guidelines [22].

Submitted: December 15, 2025 **Accepted:** February 10, 2026 **Published Online:** February 17, 2026

How to cite this article: Özkara M, Çimen Ş, Hıdıroğlu MM. Not All That Masses Are Cancer: Pilomatrixoma Presenting as Breast Cancer. *Eur Res J.* 2026;12(7):798-802. doi: 10.18621/eurj.1842449

Corresponding author: Şebnem Çimen, MD., Phone: +90 312 844 30 10, E-mail: sebnemay93@gmail.com

Available Online at https://www.eurj.org.tr

CASE PRESENTATION

Case 1

A 40-year-old male patient presented to our outpatient clinic with a 5×4 cm lump, located 3 cm from the areola in the upper outer quadrant of the right breast. The lump was accompanied by discharge and superficial infection and was raised above the skin surface (Figure 1). The patient had no comorbid conditions, history of surgery or family history of cancer. On physical examination, the mass lesion was found to have sharp margins and was mobile, while an ultrasonographic examination revealed a mass lesion with a marked posterior acoustic shadow, calcified foci, and hypervascularization (BI-RADS 5). Complete blood count, liver function tests, and kidney function tests showed normal findings, and the patient's C-reactive protein (CRP) level was within the normal range. The patient was referred to the interventional radiology unit for tru-cut biopsy and pathological sampling. The pathological examination revealed a pilomatrixoma, and the patient underwent elective surgery after surgical preparation. During surgery, the mass lesion was resected with an ellipsoid incision, circumcising the lesion, and the surgical material was sent to the pathology laboratory for examination. A Jackson-Pratt drain was placed in the resection site, and skin and subcutaneous tissues were closed as per the routine breast surgery procedure. After a 2-day follow-up period at our clinic, the amount of drainage decreased to less than 20 cc/day. The patient was discharged with full recovery after the removal of the drain. Consistent with the result of the tru-cut biopsy examination, the pathological examination of surgical material also revealed pilomatrixoma with a diameter of 4.4 cm (Malherbe's tumor) (Figure 3).

Case 2

A 70-year-old male patient presented to our center with a 4×4 cm lump, located 2 cm from the areola in the lower outer quadrant of the left breast (Figure 2). The patient had a history of diabetes mellitus but no history of surgery or family history of cancer. Physical examination revealed a firm and mobile mass lesion with sharp margins below the skin. Complete blood count, liver function tests, and kidney function tests

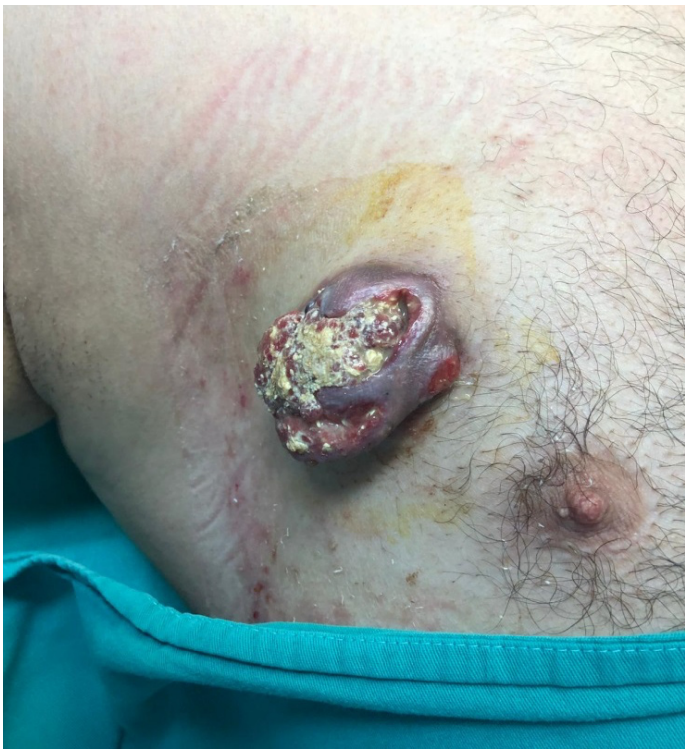

FIGURE 1. A 40-year-old male patient presented to our outpatient clinic with an infected lesion in the upper outer quadrant of the right breast.

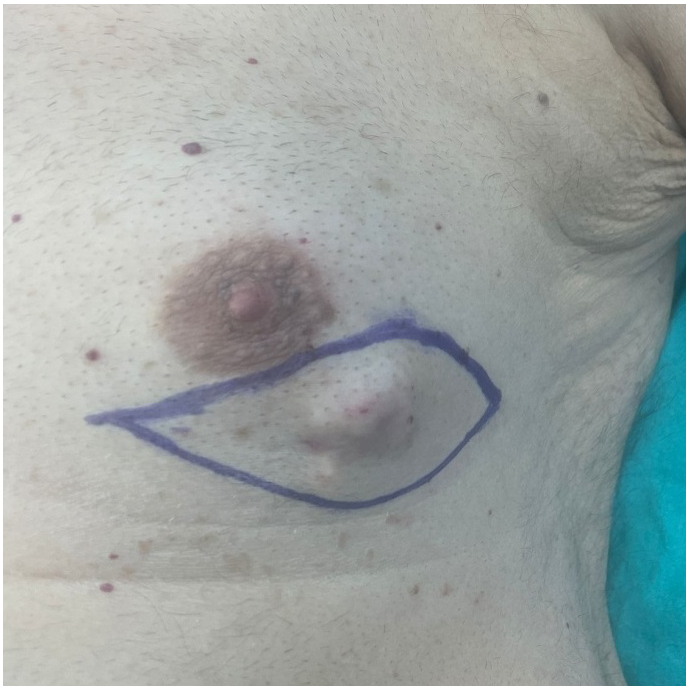

FIGURE 2. A 70-year-old male patient presented to our outpatient clinic with a palpable mass in the lower outer quadrant of the left breast.

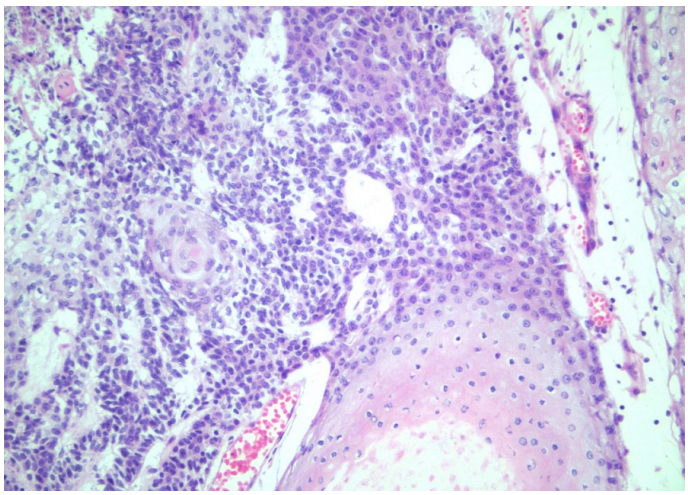

FIGURE 3. **First patient, pathological appearance (×200 magnification)**

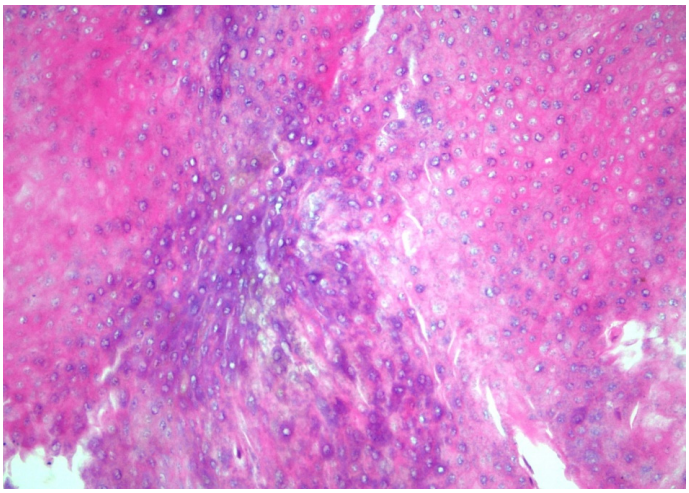

FIGURE 4. **Second patient, pathological appearance (×200 magnification)**

showed normal findings, and the patient's CRP level was within the normal range. Ultrasonographic examination revealed dense calcification within the lesion, and as the posterior of the mass could not be assessed, further investigation was recommended (BI-RADS 4), for which the patient was referred to the interventional radiology unit for tru-cut biopsy, the pathological examination of which revealed a skin tag tumor consistent with pilomatrixoma. The patient underwent elective surgery after surgical preparation. The mass lesion was resected with an ellipsoid incision, the surgical material was sent to the pathology laboratory, and the skin and subcutaneous tissues were closed. The patient was followed up for two days after surgery and discharged with full recovery. The pathological examination of the surgical material confirmed the diagnosis of pilomatrixoma (Figure 4).

No lesion was detected in either patient during control visits at 3, 6, and 12 months.

Written informed consent was obtained from both patients for publication of their clinical data and accompanying images.

DISCUSSION

Pilomatrixoma was described for the first time by Dr. Chenantals Malherbe in 1880 as a calcified epithelioma of sebaceous glands [11]. This benign lesion most often affects females in the first and second decades of life [8]. The usual localizations are the head and neck (64%), upper extremities (22%), trunk (8%) and lower extremities (5%) [21]. The occurrence of such lesions in the breast tissue of males is extremely rare [17]. Although the underlying etiology is unknown, repeated trauma and inflammation are believed to stimulate the hair follicle matrix [17]. The patients may present with a palpable mass on physical examination or with an infected lesion, as in the first case, and the lesion can mimic breast cancer. Lesions can appear hypoechoic, isoechoic or hyperechoic on ultrasonographic examination, and may contain calcifications. The most commonly reported ultrasonographic feature is posterior acoustic shadow [11, 17]. Lesions can be graded as BI-RADS 4–5 requiring further histopathological examination. The cytological features of pilomatrixoma were published for the first time by Woyke *et al*. [23] in 1982. On histopathological examination, pilomatrixoma is characterized by epithelial cells with peripheral inflammatory-like elements and nodular aggregates of connective tissue matrix. Each nodule contains two different types of epithelial cells: packs of viable, basophilic, keratin-producing cells in the periphery, and non-viable eosinophilic cells with open spaces in the center that are described as ghosts or shadows. Lesions may also contain hair, calcifications, foci of necrosis, and giant cells with multiple nuclei [15].

According to the study by Woyke *et al*. [23], pathological diagnosis can be established from a meticulous examination of a fine needle aspiration biopsy, although reports in the literature recommend tru-cut biopsy. The differentiation of the lesion from breast cancer is crucial in the practice of general surgery. The resection of the mass lesion with clear margins was found to be adequate in the two patients reported in this manuscript [21].

CONCLUSION

Pilomatrixoma is a benign skin lesion taking the form of a calcified epithelioma of the sebaceous glands, the occurrence of which in male breast tissue is extremely rare. The lesion may be confused with breast cancer based on clinical and ultrasonographic findings, and the use of tru-cut biopsy is recommended in the differential diagnosis. The resection of the lesion with clear margins would seem to be sufficient as a surgical treatment.

Ethics Approval and Consent to Participate

This study is a case report and therefore did not require approval from an institutional ethics committee. Patient data were used retrospectively. Written informed consent was obtained from both patients for the use of their clinical information and accompanying images for scientific publication purposes.

Data Availability

All data generated or analyzed during this study are included in this published article. The data that support the findings of this study are available on request from the corresponding author, upon reasonable request.

Authors' Contribution

Study Conception: MÖ, ŞÇ; Study Design: ŞÇ, MMH; Supervision: MÖ, ŞÇ; Funding: MÖ; Materials: MMH, ŞÇ; Data Collection and/or Processing: MMH, ŞÇ; Statistical Analysis and/or Data Interpretation: ŞÇ, MMH; Literature Review: MÖ, ŞÇ; Writer: MÖ, ŞÇ; and Critical Review: MMH.

Conflict of Interest

The author(s) disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The author(s) disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgments

The authors have no acknowledgments to declare.

Generative Artificial Intelligence Statement

The author(s) declare that no artificial intelligence-based tools or applications were used during the preparation process of this manuscript. The all content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

Editor's Note

All statements made in this article are solely those of the authors and do not represent the views of their affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Hubeny CM, Sykes JB, O'Connell A, Dogra VS. Pilomatrixoma of the adult male breast: a rare tumor with typical ultrasound features. J Clin Imaging Sci. 2011;1:12. doi: 10.4103/2156-7514.76690.
2. Rousselot C, Tourasse C, Samimi M, Degand P, Dénier JF, Michenet P. Pilomatricomes mammaires révélés par des microcalcifications à la mammographie: à propos de deux cas [Breast pilomatrixoma manifested as microcalcifications on mammography: report of two cases]. J Radiol. 2007;88(7-8 Pt 1):978-80. doi: 10.1016/s0221-0363(07)89907-0. [Article in French]
3. Devi K, Gupta AK, Saigal RK. Pilomatrixoma of Male Breast. Indian J Dermatol Venereol Leprol. 1982;48(2):102-104.
4. Hamilton A, Young GI, Davis RI. Pilomatrixoma mimicking breast carcinoma. Br J Dermatol. 1987;116(4):58558-6. doi: 10.1111/j.1365-2133.1987.tb05883.x.
5. Gilles R, Guinebretière JM, Gallay X, Vanel D. Pilomatrixoma mimicking male breast carcinoma on mammography. AJR Am J Roentgenol. 1993;160(4):895. doi: 10.2214/ajr.160.4.8456689.
6. Ismail W, Pain S, al-Okati D, al Sewan M. Giant pilomatricoma simulating carcinoma of the male breast. Int J Clin Pract. 2000 ;54(1):55-56.
7. Imperiale A, Calabrese M, Monetti F, Zandrino F. Calcified pilomatrixoma of the breast: mammographic and sonographic findings. Eur Radiol. 2001;11(12):2465-2467. doi: 10.1007/s003300000798.
8. Ali MZ, Ali FZ. Pilomatrixoma breast mimicking carcinoma.

J Coll Physicians Surg Pak. 2005;15(4):248-9.
9. Becker TS, Moreira MA, Lima LA, de Oliveira EL, Freitas-Júnior R. Pilomatrixoma mimicking breast cancer in man. Breast J. 2010 Jan-Feb;16(1):89-91. doi: 10.1111/j.1524-4741.2009.00852.x.
10. AlSharif S, Meguerditchian A, Omeroglu A, Lamarre P, Altinel G, Mesurolle B. Pilomatricoma of the male breast: sonographic mammographic MRI features with pathologic correlation. Clin Imaging. 2015;39(2):308-310. doi: 10.1016/j.clinimag.2014.07.005.
11. Nori J, Abdulcadir D, Giannotti E, Calabrese M. Pilomatrixoma of the breast, a rare lesion simulating breast cancer: a case report. J Radiol Case Rep. 2013;7(10):43-50. doi: 10.3941/jrcr.v7i10.1651.
12. Käppeli M, Routiot T, Judlin P. Un diagnostic différentiel du cancer du sein chez l'homme: l'épithélioma momifié de Malherbe, ou Pilomatricome [A male breast cancer differential diagnosis: Calcific epithelioma of Malherbe or Pilomatrixoma]. Gynecol Obstet Fertil. 2016;44(3):190-191. doi: 10.1016/j.gyobfe.2016.02.002. [Article in French]
13. Fama' F, Ieni A, Tchernev G, et al. Pilomatrixoma of the breast in a patient with type 1 myotonic dystrophy: successful surgical approach. J Biol Regul Homeost Agents. 2016;30(2 Suppl 2):1-6.
14. Kapoor A, Narayanan R, Tandon A, Santosh AK. Pilomatricoma: An unusual cause of lump in a male breast. J Clin Ultrasound. 2018;46(3):209-211. doi: 10.1002/jcu.22503.
15. Pai T, Harwani SR, Patil A, et al. Pilomatrix Carcinoma Masquerading as Breast Carcinoma. Indian J Med Paediatr Oncol. 2017;38(3):367-370. doi: 10.4103/ijmpo.ijmpo_118_16.
16. Ward RC, Donegan L, Khalil H, Wang Y. Pilomatricoma of the male breast. Breast J. 2019;25(5):1012-1013. doi: 10.1111/tbj.13406.
17. Clark A, Leddy R, Spruill L, Cluver A. Pilomatrixoma, a Rare Mimicker of Male Breast Cancer. J Clin Imaging Sci. 2019;9:46. doi: 10.25259/JCIS_64_2019.
18. Sood N, Raj B. Pilomatricoma male breast, mimicking breast carcinoma-A rare case. Indian J Pathol Microbiol. 2021;64(1):204-205. doi: 10.4103/IJPM.IJPM_194_20.
19. Bensalah A, Benaaddach HO, Gouzi I, et al. Pilomatrixoma mimicking a breast neoplasm: imaging finding in an uncommon case report. Radiol Case Rep. 2021;16(9):2357-2361. doi: 10.1016/j.radcr.2021.06.007.
20. Al-Issai M, Al-Rahbi S, Al-Futaisi A, Gokhale UA, Al-Salhi S. Pilomatrixoma of the Male Breast: Case Report and Literature Review. Oman Med J. 2023;38(4):e533. doi: 10.5001/omj.2023.30.
21. Jones CD, Ho W, Robertson BF, Gunn E, Morley S. Pilomatrixoma: A Comprehensive Review of the Literature. Am J Dermatopathol. 2018;40(9):631-641. doi: 10.1097/DAD.0000000000001118.
22. Agha RA, Franchi T, Sohrabi C, Mathew G, Kerwan A; SCARE Group. The SCARE 2020 Guideline: Updating Consensus Surgical CAse REport (SCARE) Guidelines. Int J Surg. 2020;84:226-230. doi: 10.1016/j.ijsu.2020.10.034.
23. Woyke S, Olszewski W, Eichelkraut A. Pilomatrixoma: a pitfall in the aspiration cytology of skin tumors. Acta Cytol. 1982;26(2):189-194.