The Medical Bulletin of Haseki

E-ISSN: 2147-2688

2026
Volume 64
Issue
September 4

www.hasekidergisi.com

The Medical Bulletin of Haseki

Editorial Board

Publisher Contact
Address: Molla Gurani Mah. Kacamak Sk. No: 21/1 34093 Istanbul, Türkiye
Phone: +90 (530) 177 30 97 / +90 (539) 307 32 03
E-mail: info@galenos.com.tr/yayin@galenos.com.tr
Web: www.galenos.com.tr
Publisher Certificate Number: 14521

Online Publishing Date: September 2026
E-ISSN: 2147-2688
International scientific journal published quarterly.

The Medical Bulletin of Haseki

Scientific Advisory Board

The Medical Bulletin of Haseki

Scientific Advisory Board

The Medical Bulletin of Haseki

Scientific Advisory Board

The Medical Bulletin of Haseki

Please refer to the journal's webpage **(https://www.hasekidergisi.com/**) for "About Us", "Instructions to Authors" and "Peer Review & Ethics".

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

The Medical Bulletin of Haseki is indexed in **Emerging Sources Citation Index (ESCI), EBSCO Database, Gale, Turkish Medline-National Citation Index, Excerpta Medica/EMBASE, SCOPUS, TÜBİTAK/ULAKBİM, CINAHL, DOAJ, Hinari, GOALI, ARDI, OARE, AGORA, ProQuest, J-Gate, IdealOnline** and **Türkiye Citation Index**.

Owner: Mehmet Mesut Sonmez on University of Health Sciences Türkiye, İstanbul Haseki Training and Research Hospital

Responsible Manager: Akif ERBİN

The Medical Bulletin of Haseki

Contents

Original Articles

Original Article

DOI: 10.4274/haseki.galenos.2026.63935
Med Bull Haseki 2026;64(4):222-228

Association of Circulating FGF-23 and Klotho with Cardiovascular Disease in Patients with Chronic Kidney Disease

Ayla Yildiz[1], Levent Deniz[2]

[1]*University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital, Clinic of Medical Biochemistry, Istanbul, Türkiye*
[2]*University of Health Sciences Türkiye, Istanbul Training and Research Hospital, Clinic of Medical Biochemistry, Istanbul, Türkiye*

Abstract

Aim: Chronic kidney disease (CKD) contributes to an increased risk of cardiovascular disease (CVD) beyond traditional risk factors, partly through the fibroblast growth factor 23 (FGF-23) and Klotho pathways. This study examined the association between circulating FGF-23 and Klotho levels and CVD in individuals with CKD.

Methods: This cross-sectional study included 110 patients with CKD treated at University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital. Patients were classified into CVD-positive (n=56) and CVD-negative (n=54) groups according to the presence of CVD, defined as coronary artery disease, ischaemic stroke, or peripheral arterial disease. Serum FGF-23 and Klotho levels were quantified using enzyme-linked immunosorbent assay, and routine biochemical and hormonal parameters were simultaneously assessed. Statistical analyses included group comparisons, correlation analyses, and binary logistic regression.

Results: Patients with CVD had higher serum FGF-23 levels than those without CVD (p=0.022), whereas Klotho levels did not differ between the groups. Multivariate regression analyses showed that FGF-23 was associated with CVD [odds ratio =1.0026 (1.0002-1.0049), p=0.032] after adjusting for estimated glomerular filtration rate, age, sex, diabetes mellitus, and hypertension. Klotho showed inverse correlations with creatinine (r=-0.297, p=0.029) in CVD-negative patients and with glucose (r=-0.299, p=0.025) in CVD-positive patients and was positively associated with FGF-23 (r=0.558, p<0.001) in CVD-positive patients.

Conclusion: Elevated FGF-23 levels were associated with CVD in patients with CKD, whereas Klotho levels did not distinguish patients with CVD. These findings suggest a possible relationship between FGF-23 and CVD in CKD; however, further studies are required.

Keywords: Cardiovascular disease, creatinine, enzyme-linked immunosorbent assay, fibroblast growth factor, renal insufficiency

Introduction

Chronic kidney disease (CKD) is a major global health problem and is projected to become the fifth leading cause of years of life lost by 2040, with cardiovascular disease (CVD) accounting for over half of the deaths in this population (1,2). The excess cardiovascular risk observed in patients with CKD cannot be fully explained by traditional risk factors, suggesting the involvement of disease-specific mechanisms (2,3). Disturbances in mineral metabolism are a central feature of CKD and play a critical role in the development of cardiovascular complications. In this context, the fibroblast growth factor 23 (FGF-23)-Klotho axis has emerged as a key pathway linking kidney dysfunction and cardiovascular pathology (4,5).

Fibroblast growth factor 23 levels increase early in the course of CKD, often preceding abnormalities in phosphate or parathyroid hormone levels, whereas Klotho expression progressively declines as kidney function deteriorates (5,6). Compared to Klotho, FGF-23 has been more extensively investigated, and accumulating evidence has linked elevated FGF-23 levels to adverse cardiovascular outcomes in CKD, particularly cardiac remodelling and left ventricular hypertrophy (7). Despite increasing evidence

Corresponding Author: Ayla Yildiz, University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital, Clinic of Medical Biochemistry, Istanbul, Türkiye
E-mail: aylayildizm@gmail.com **ORCID:** orcid.org/0000-0002-2982-9918
Received: 14.01.2026 **Accepted:** 24.03.2026 **Publication Date:** 24.09.2026

Cite this article as: Yildiz A, Deniz L. Association of circulating FGF-23 and Klotho with cardiovascular disease in patients with chronic kidney disease. Med Bull Haseki.2026;64(4):222-228

supporting the individual roles of FGF-23 and Klotho, data evaluating their combined association with established CVD in patients with CKD remain limited (8).

We hypothesised that circulating FGF-23 and Klotho levels are associated with CVD in patients with CKD and may serve as independent cardiovascular risk markers beyond traditional risk factors. This study aimed to investigate the relationship between FGF-23, Klotho, and CVD in patients with CKD.

Materials and Methods

Compliance with Ethical Standards

The study was conducted with approval of the University of Health Sciences Türkiye, Istanbul Training and Research Hospital Clinical Research Ethics Committee (approval number: 1448, date: 28.09.2018). Any personal or identifiable information was anonymised throughout the study to ensure confidentiality. The study was carried out in line with the ethical standards outlined in the Helsinki Declaration. Written informed consent was obtained from all the participants before their inclusion in the study.

Study Design and Patient Selection

We analysed 110 consecutive patients with CKD who were admitted to the Nephrology Clinic of University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital between October 2018 and April 2019. In all cases, serum FGF-23 and Klotho levels were assessed simultaneously with routine biochemical and hormonal tests, such as creatinine, estimated glomerular filtration rate (eGFR), urea, calcium, magnesium, phosphorus, parathyroid hormone, glucose, and 25-hydroxyvitamin D. The eGFR was calculated using the CKD Epidemiology Collaboration equation. Clinical data for each patient, including cardiovascular status, smoking history, hypertension, diabetes mellitus, coronary artery disease, ischaemic stroke, and peripheral artery disease, were recorded. Cardiovascular disease was defined as the presence of coronary artery disease, ischaemic stroke, or peripheral arterial disease. Coronary artery disease was defined as a history of percutaneous coronary intervention (stent implantation) or coronary artery bypass grafting. Ischemic stroke was defined based on a diagnosis documented in the medical history and supported by imaging findings. Peripheral arterial disease was defined as a history of peripheral revascularisation or imaging-confirmed disease.

Laboratory Measurements

Routine biochemical and hormonal analyses were performed using an AU5800 analyser (Beckman Coulter, Brea, CA, USA) and a Siemens Immulite 1000 Immunoassay System (Siemens Healthcare Diagnostics, Los Angeles, CA, USA), respectively. Commercial enzyme-linked immunosorbent assay kits were used to assess serum FGF-23 and Klotho levels following the manufacturer's guidelines (Sunred Biological Technology Co. Ltd., Shanghai, China). The analytical sensitivity [limit of detection (LOD)] and measurement ranges of the assays were defined according to the manufacturer's specifications (catalogue numbers: FGF-23, 201-12-0060; Klotho, 201-12-2782). The LOD for FGF-23 was 5.147 pg/mL, with a measurement range of 10-1500 pg/mL. The LOD for Klotho was 0.05 ng/mL, with a measurement range of 0.1-20 ng/mL. The within-run coefficients of variation (CVs) for FGF-23 and Klotho were <10%, and the inter-assay CVs were <12%.

Individuals aged 18 years and above who were diagnosed with CKD, who were followed at the Nephrology Clinic of University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, and who had available serum FGF-23 and Klotho measurements and complete cardiovascular clinical data were included. Patients younger than 18 years; those diagnosed with acute kidney injury; those receiving kidney replacement therapy; those with a history of kidney transplantation; and those with active infection, inflammatory or autoimmune disease, malignancy, chronic liver disease, or an acute cardiovascular event within the previous three months were excluded (Figure 1).

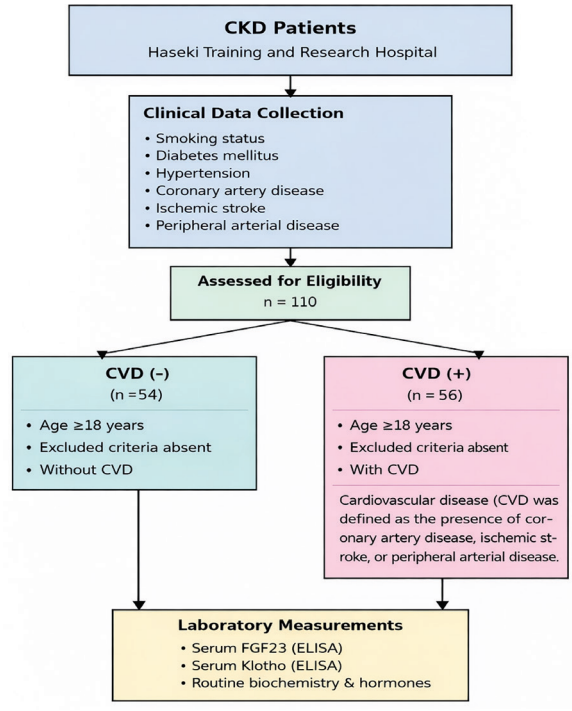

Figure 1. Flow diagram of the study population and participant selection

CKD: Chronic kidney disease, CVD: Cardiovascular disease, ELISA: Enzyme-linked immunosorbent assay, FGF-23: Fibroblast growth factor 23

Sample Size Estimation

The sample size was estimated a priori based on the expected difference in serum FGF-23 levels between patients with and without CVD. Using an effect size of Cohen's d=1.19 derived from the findings of Khodeir et al. (5), a two-sided alpha level of 0.05 and 95% power, the minimum required total sample size was calculated as 40 participants. To improve the precision of the estimates and account for potential exclusions or incomplete data, all 110 eligible patients were included in the analysis. The sample size calculation was performed using G*Power version 3.1.9.7 software (Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany).

Statistical Analysis

Statistical analyses were carried out using IBM SPSS v. 26 (Armonk, NY, USA) and GraphPad Prism v. 10.4.2 (San Diego, CA, USA). Normality of the data was assessed using the Kolmogorov-Smirnov test. Categorical variables were examined using the chi-square test or Fisher's exact test. Student's t-test was used for parametric data, and the Mann-Whitney U test was used for non-parametric data. The Kruskal-Wallis test was used to compare FGF-23 and Klotho levels among the three independent groups. Spearman's rank correlation analysis was conducted to explore the relationships between FGF-23, Klotho, and other variables. In the binary logistic regression model, the FGF-23 level was included as an independent variable, and the analysis was adjusted for eGFR, age, sex, diabetes mellitus, and hypertension to assess its association with the presence of CVD. Covariates were selected a priori based on their established clinical relevance and evidence from previous literature. Statistical significance was defined as $p<0.05$.

Results

Table 1 presents the demographic features of the CVD-positive and CVD-negative groups. The groups did not differ significantly with respect to age, sex, smoking status, or hypertension prevalence ($p>0.05$). However, the CVD-positive group exhibited significantly higher incidence rates of diabetes mellitus ($p=0.022$), coronary artery disease ($p<0.001$), ischaemic stroke ($p=0.027$), and peripheral arterial disease ($p=0.027$) compared with the CVD-negative group (Table 1).

Table 2 presents the results of the between-group statistical analyses of FGF-23, Klotho, and other biochemical variables evaluated in the study. Fibroblast growth factor 23 levels were elevated in the CVD-positive group [130 (38.7-231)] compared to the CVD-negative group [69.9 (33.6-141)], with a p-value of 0.022 (Figure 2). The groups did not differ significantly with respect to levels of Klotho, creatinine, urea, calcium, phosphorus, magnesium, parathyroid hormone, and 25-hydroxyvitamin D. Additionally, glucose levels were higher in the CVD-positive group [139 (109-186)] than in the CVD-negative group [108 (96.0-136)]; $p=0.002$ (Table 2).

Table 1. Baseline characteristics of the CVD (+) and CVD (-) groups

Variables	CVD (-) (n=54)	CVD (+) (n=56)	p-value
Age (years), mean ± SD	64.2±9.02	67.4±8.24	0.054[a]
Gender, n (%) Female Male	 23 (42.6) 31 (57.4)	 19 (33.9) 37 (66.1)	0.350[b]
Smoking status, n (%) No Yes	 36 (66.7) 18 (33.3)	 36 (64.3) 20 (35.7)	0.951[b]
Diabetes mellitus, n (%) No Yes	 29 (53.7) 25 (46.3)	 17 (30.4) 39 (69.6)	**0.022[b]**
Hypertension, n (%) No Yes	 7 (13.0) 47 (87.0)	 5 (8.90) 51 (91.1)	0.709[b]
Coronary artery disease, n (%) No Yes	 54 (100) 0 (0.00)	 4 (7.10) 52 (92.9)	**<0.001[b]**
Ischemic stroke, n (%) No Yes	 54 (100) 0 (0.00)	 50 (89.3) 6 (10.7)	**0.027[b]**
Peripheral arterial disease, n (%) No Yes	 54 (100) 0 (0.00)	 50 (89.3) 6 (10.7)	**0.027[b]**

Statistically significant values ($p<0.05$) are shown in bold. [a]Student's t-test, [b]Chi-square
CVD: Cardiovascular disease, SD: Standard deviation

Table 2. Laboratory parameters of the CVD (+) and CVD (-) groups			
Variables	**CVD (-) (n=54)**	**CVD (+) (n=56)**	**p-value**
FGF-23 (pg/mL), median (25^{th}-75^{th} percentiles)	69.9 (33.6-141)	130 (38.7-231)	**0.022[a]**
Klotho (ng/mL), median (25^{th}-75^{th} percentiles)	4.00 (3.34-4.90)	4.29 (3.54-6.23)	0.149[a]
Creatinine (mg/dL), median (25^{th}-75^{th} percentiles)	1.80 (1.46-2.60)	1.89 (1.41-2.54)	0.902[a]
eGFR (mL/min/1.73 m^2), mean ± SD	32.4±12.9	34.4±14.5	0.450[b]
Urea (mg/dL), median (25^{th}-75^{th} percentiles)	77.7 (58.0-94.5)	68.2 (49.0-106)	0.403[a]
Calcium (mg/dL), median (25^{th}-75^{th} percentiles)	9.70 (9.40-10.0)	9.60 (9.10-9.95)	0.410[a]
Phosphorus (mg/dL), mean ± SD	3.87±0.63	3.76±0.92	0.441[b]
Magnesium (mg/dL), mean ± SD	2.11±0.38	2.06±0.35	0.424[b]
Glucose (mg/dL), median (25^{th}-75^{th} percentiles)	108 (96.0-136)	139 (109-186)	**0.002[a]**
25(OH)D (ng/mL), median (25^{th}-75^{th} percentiles)	16.9 (11.6-24.8)	14.8 (10.3-24.2)	0.453[a]
PTH (pg/mL), median (25^{th}-75^{th} percentiles)	94.0 (67.0-149)	122 (61.0-188)	0.212[a]

Statistically significant values (p<0.05) are shown in bold. Variables with a normal distribution were expressed as mean ± standard deviation, while variables without a normal distribution were expressed as median (25^{th}-75^{th} percentiles). [a]Mann-Whitney U test, [b]Student's t-test
FGF-23: Fibroblast growth factor 23, eGFR: Estimated glomerular filtration rate, 25(OH)D: 25-hydroxyvitamin D, PTH: Parathyroid hormone, CVD: Cardiovascular disease, SD: Standard deviation

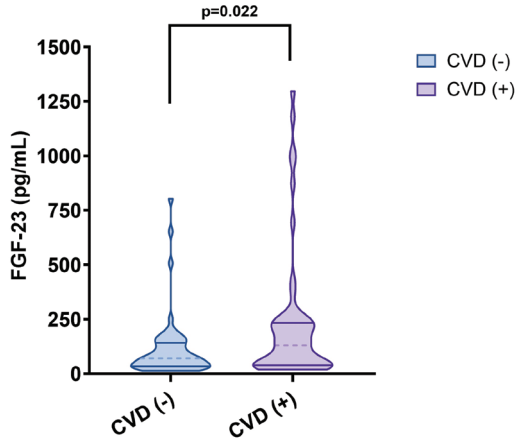

Figure 2. Comparison of FGF-23 levels in patients categorized by cardiovascular disease status
CVD: Cardiovascular disease, FGF-23: Fibroblast growth factor 23

In the univariate regression analysis, FGF-23 demonstrated a significant association with the presence of CVD [odds ratio (OR)=1.0024 (1.0002-1.0047), p=0.035]. In the multivariate regression model, FGF-23 was associated with CVD [OR=1.0026 (1.0002-1.0049), p=0.032] after adjusting for eGFR, age, sex, diabetes mellitus, and hypertension (Table 3).

Fibroblast growth factor 23 levels were positively correlated with Klotho levels in the CVD-positive group (r=0.558, p<0.001). Conversely, Klotho levels were negatively correlated with creatinine levels (r=-0.297, p=0.029) in the CVD-negative group. Additionally, Klotho levels were negatively correlated with glucose levels (r=-0.299, p=0.025) in the CVD-positive group (Table 4).

Patients were categorised into three groups based on eGFR levels, corresponding to CKD stages: 45-59 (n=27), 30-44 (n=35), and <30 (n=48) mL/min/1.73 m^2. Fibroblast growth factor 23 levels were not significantly different among the groups (p=0.496), with median (25^{th}-75^{th} percentile) values of 71.9 (25.3-167), 90.9 (57.5-185), and 106 (34.2-191), respectively. Similarly, Klotho levels did not differ significantly among the groups (p=0.999); median (25^{th}-75^{th} percentiles) values were 4.28 (3.30-5.82), 4.28 (3.29-6.34), and 4.11 (3.60-5.58), respectively.

Discussion

This study assessed the relationship of circulating FGF-23 and Klotho with the occurrence of CVD in patients with CKD. Our findings indicated that FGF-23 was significantly higher in the CVD-positive group and remained independently associated with CVD after controlling for traditional cardiovascular risk factors. Conversely, serum Klotho levels did not differ significantly between patients with and without CVD. These findings are consistent with previous studies suggesting an association between FGF-23 and cardiovascular risk (9-11).

The study groups were demographically similar; no notable differences in age, sex, smoking habits, or hypertension were observed between groups. The higher prevalence of diabetes mellitus and atherosclerotic cardiovascular conditions in patients with CVD is consistent with epidemiological data in CKD populations and aligns with previously reported findings (2). The similar distribution of traditional risk factors supports the notion that the observed association between FGF-23 and CVD is not solely attributable to baseline clinical differences.

Table 3. Binary logistic regression analysis results for effects of FGF-23 on cardiovascular disease				
Variables	**Univariate**		**Multivariate***	
	OR (95% CI)	**p-value**	**OR (95% CI)**	**p-value**
FGF-23 (pg/mL)	1.0024 (1.0002-1.0047)	**0.035**	1.0026 (1.0002-1.0049)	**0.032**

Statistically significant values (p<0.05) are shown in bold
*Adjusted for eGFR, hypertension, diabetes mellitus, age, and sex.
CI: Confidence interval, OR: Odds ratio, eGFR: Estimated glomerular filtration rate, FGF-23: Fibroblast growth factor 23

Table 4. Correlations between FGF-23 and Klotho and other variables								
Parameter	**FGF-23 (pg/mL)**				**Klotho (ng/mL)**			
	CVD (-) (n=54)		**CVD (+) (n=56)**		**CVD (-) (n=54)**		**CVD (+) (n=56)**	
	r	**p-value**	**r**	**p-value**	**r**	**p-value**	**r**	**p-value**
Klotho (ng/mL)	0.248	0.071	**0.558**	**<0.001**	1.000	<0.001	1.000	<0.001
Creatinine (mg/dL)	0.076	0.583	-0.063	0.643	**-0.297**	**0.029**	0.111	0.413
eGFR (mL/min/1.73 m²)	-0.123	0.377	0.049	0.722	0.216	0.117	-0.136	0.319
Urea (mg/dL)	0.054	0.699	-0.039	0.774	-0.162	0.243	0.111	0.417
Calcium (mg/dL)	-0.027	0.844	-0.127	0.349	-0.001	0.995	-0.088	0.520
Phosphorus (mg/dL)	-0.133	0.337	-0.049	0.718	0.093	0.502	0.114	0.404
Magnesium (mg/dL)	0.112	0.421	0.101	0.458	-0.098	0.480	0.127	0.353
Glucose (mg/dL)	-0.106	0.446	-0.191	0.157	-0.115	0.406	**-0.299**	**0.025**
25(OH)D (ng/mL)	0.155	0.264	0.026	0.848	-0.090	0.518	-0.157	0.249
PTH (pg/mL)	0.011	0.935	-0.172	0.206	-0.039	0.777	-0.054	0.695

Statistically significant values (p<0.05) are shown in bold
FGF-23: Fibroblast growth factor 23, eGFR: Estimated glomerular filtration rate, 25(OH)D: 25-hydroxyvitamin D, PTH: Parathyroid hormone, CVD: Cardiovascular disease

Elevated FGF-23 levels in patients with CVD represent a central finding of this study. This observation is in agreement with large cohort studies and meta-analyses demonstrating that increased FGF-23 concentrations are associated with heart failure, left ventricular hypertrophy, cardiovascular events, and mortality across different CKD stages (10,11).

The groups did not differ significantly in kidney function parameters and traditional markers of mineral metabolism. This finding indicates that the association between FGF-23 and CVD cannot be attributed solely to advanced renal dysfunction or pronounced disturbances in mineral metabolism. Previous studies have similarly demonstrated that FGF-23 is associated with cardiovascular risk independently of serum phosphate and parathyroid hormone levels (9,10).

In contrast to FGF-23 levels, serum Klotho levels were not significantly different between patients with and without CVD. This observation aligns with findings from cross-sectional studies highlighting the limited discriminatory capacity of Klotho for established CVD. Data from the chronic renal insufficiency cohort study (12) indicated that circulating Klotho levels were not correlated with mortality, hospitalisation due to heart failure, atherosclerotic cardiovascular events, or the progression of kidney disease in patients with CKD. Furthermore, Klotho deficiency did not confound or alter the association between FGF-23 and adverse clinical outcomes. These results imply that circulating Klotho may play a limited role in determining short- to mid-term clinical outcomes in CKD (12). Consistent with these findings, serum Klotho levels in the present study did not differ between patients with and without established CVD and did not independently discriminate CVD status. In contrast, FGF-23 levels appeared to be independently associated with CVD. Overall, these results show that FGF-23 is more closely linked to cardiovascular risk than circulating Klotho levels in people with CKD.

Multivariate analysis showed an association between FGF-23 and CVD after adjusting for eGFR, age, sex, diabetes mellitus, and hypertension. Although the effect size per unit increase appeared modest, the wide biological range of FGF-23 levels observed in patients with CKD suggests that cumulative increases may have clinically meaningful implications. Previous studies have reported dose-dependent and nonlinear associations between FGF-23 levels and adverse cardiovascular outcomes, particularly at high concentrations (10,13).

In this study, the moderate and statistically significant positive correlation observed between FGF-23 and Klotho in the CVD-positive group (r=0.558, p<0.001) suggests

that the FGF-23–Klotho axis may exhibit a different regulatory pattern in the presence of CVD. Although FGF-23 and Klotho are classically considered to have an inverse physiological relationship, previous studies have reported inconsistent associations between these biomarkers in CKD (7,12). In addition, circulating soluble Klotho does not necessarily reflect tissue-level expression or biological activity (14,15), which may partly explain the unexpected findings. Furthermore, both FGF-23 and Klotho may be influenced by shared pathophysiological processes in CKD and CVD, potentially leading to parallel changes under certain conditions (8). Therefore, the positive correlation observed in our study may reflect disease-specific regulatory mechanisms rather than direct physiological interactions. The absence of a significant association in the CVD-negative group indicates that this parallel increase may be specific to the disease context. Additionally, the inverse relationships observed between Klotho levels and creatinine and glucose levels support previous findings linking reduced Klotho availability to declining renal function and metabolic dysregulation (7,12). Although FGF-23 and Klotho are known to constitute a closely related physiological axis, the observed correlation should not be interpreted as evidence of a causal relationship and may be influenced by concomitant renal function, metabolic status, or other regulatory factors.

In this context, the positive correlation observed between FGF-23 and Klotho in the CVD-positive group may reflect a combination of compensatory Klotho upregulation, functional Klotho resistance, chronic cardiovascular stress, increased Klotho shedding, and disease-specific regulatory mechanisms. This finding does not indicate preservation of the physiological FGF-23-Klotho balance; rather, it emphasises the complex nature of this axis in established CVD and highlights the limitations of interpreting circulating biomarkers in isolation (16-19).

Although the per-unit effect size of FGF-23 was modest, its wide biological range and previously demonstrated nonlinear and dose-dependent associations with cardiovascular outcomes suggest that even small increases may be clinically meaningful (9,20-22). As reported in previous studies, potential variability in biomarker measurements should be considered when interpreting these findings (9,12,15). Cardiovascular disease was defined by imaging and documented diagnoses, including coronary artery disease, peripheral arterial disease, and stroke; however, subclinical disease could not be excluded.

Overall, our findings are consistent with accumulating evidence that FGF-23 plays a role in CKD-associated CVD and is independently associated with cardiovascular risk beyond traditional factors (23,24). FGF-23 may represent a potential biomarker for cardiovascular risk stratification in patients with CKD; however, prospective studies are required before routine clinical application.

Study Limitations

This study has several limitations. First, its cross-sectional, single-centre design limits the ability to establish causality between FGF-23 levels and CVD and may restrict the generalisability of the findings. Therefore, the observed associations should be interpreted as correlational, and prospective longitudinal multicentre studies are needed to clarify potential causal relationships. Second, the relatively modest sample size may have limited our ability to detect subtle associations, particularly those involving serum Klotho levels. Third, CVD was identified based on clinical history and medical records, which does not permit the exclusion of subclinical disease. Furthermore, serum Klotho levels may not accurately reflect tissue-level Klotho expression or biological activity, which may partially explain the lack of an observed association between Klotho levels and CVD. Additionally, urinary albumin-to-creatinine ratio data were unavailable for all patients and were therefore excluded from the analysis, which may have limited the clinical interpretation of our findings.

Despite these limitations, this study is among recent investigations highlighting CVDs associated with CKD. In addition, the comprehensive evaluation of a wide range of biochemical parameters, together with analyses stratified by disease stage, constitutes a significant strength of the study. Moreover, the simultaneous evaluation of FGF-23 and Klotho in this patient population constitutes an additional strength, as studies assessing both biomarkers together in this group are limited.

Conclusion

Based on these findings, this study demonstrates that circulating FGF-23 levels are significantly associated with CVD in individuals with CKD, whereas serum Klotho levels do not differ by CVD status. Our findings indicate a possible association between FGF-23 and cardiovascular risk in CKD, which requires further confirmation.

Ethics

Ethics Committee Approval: The study was conducted with the approval University of Health Sciences Türkiye, Istanbul Training and Research Hospital Clinical Research Ethics Committee (approval number: 1448, date: 28.09.2018).

Informed Consent: Written informed consent was obtained from all the participants before their inclusion in the study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: A.Y., L.D., Concept: A.Y., L.D., Design: A.Y., L.D., Data Collection or Processing: A.Y., L.D., Analysis or Interpretation: A.Y., L.D., Literature Search: A.Y., L.D., Writing: A.Y., L.D.

Conflict of Interest: The authors declared that there were no conflicts of interest.

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

References

1. Li Y, Jiang R, Ouyang C, et al. Chronic kidney disease is associated with increased risk of sudden cardiac death. Nat Commun. 2025;16:9180.
2. Schunk SJ, Zimmermann P. Cardiovascular risk and its presentation in chronic kidney disease. J Clin Med. 2025;14:4567.
3. Kunz M, Götzinger F, Emrich I, Schwenger V, Böhm M, Mahfoud F. Cardio-renal interaction - clinical trials update 2022. Nutr Metab Cardiovasc Dis. 2022;32:2451-8.
4. Kim A, Parikh MA, Khalid H, Frishman WH, Peterson SJ. Fibroblast growth factor-23 and cardiovascular disease in patients with chronic kidney disease. Cardiol Rev. 2026.
5. Khodeir SA, Okda HI, Abdalal HM. Clinical significance of fibroblast growth factor-23 and soluble alpha Klotho in different stages of chronic kidney disease. Saudi J Kidney Dis Transpl. 2019;30:108-18.
6. Dohuim AHE, Ghazy MAEM, Nosair NAA, Hussien AK. Correlation linking fibroblast growth factor 23 in circulation (FGF-23) and serum aldosterone levels in different stages of chronic kidney disease. Egypt J Intern Med. 2025;37:46.
7. Płonka J, Olejnik A, Klus A, et al. Use of FGF-23 and sαKlotho for risk stratification in patients with acute heart failure. J Clin Med. 2025;14:860.
8. Edmonston D, Grabner A, Wolf M. FGF23 and klotho at the intersection of kidney and cardiovascular disease. Nat Rev Cardiol. 2024;21:11-24.
9. Isakova T, Xie H, Yang W, et al; Chronic Renal Insufficiency Cohort (CRIC) Study Group. Fibroblast growth factor 23 and risks of mortality and end-stage renal disease in patients with chronic kidney disease. JAMA. 2011;305:2432-9.
10. Faul C, Amaral AP, Oskouei B, et al. FGF23 induces left ventricular hypertrophy. J Clin Invest. 2011;121:4393-408.
11. Buiten MS, de Bie MK, Bouma-de Krijger A, et al. Soluble Klotho is not independently associated with cardiovascular disease in a population of dialysis patients. BMC Nephrol. 2014;15:197.
12. Edmonston D, Fuchs MAA, Burke EJ, Isakova T, Wolf M; Chronic Renal Insufficiency Cohort (CRIC) Study Investigators. Klotho and clinical outcomes in CKD: findings from the chronic renal insufficiency cohort (CRIC) study. Am J Kidney Dis. 2024;84:349-60.e1.
13. Law JP, Price AM, Pickup L, et al. Clinical potential of targeting fibroblast growth factor-23 and αKlotho in the treatment of uremic cardiomyopathy. J Am Heart Assoc. 2020;9:e016041. Erratum in: J Am Heart Assoc. 2020;9:e014566.
14. Hu MC, Shi M, Zhang J, et al. Klotho deficiency causes vascular calcification in chronic kidney disease. J Am Soc Nephrol. 2011;22:124-36.
15. Seiler S, Rogacev KS, Roth HJ, et al. Associations of FGF-23 and sKlotho with cardiovascular outcomes among patients with CKD stages 2-4. Clin J Am Soc Nephrol. 2014;9:1049-58.
16. Lu X, Hu MC. Klotho/FGF23 axis in chronic kidney disease and cardiovascular disease. Kidney Dis (Basel). 2017;3:15-23.
17. Waitupu A, Pratiwi L, Sutanto H, Santoso D, Hertanto DM. Molecular pathophysiology of chronic kidney disease-mineral and bone disorder: Focus on the fibroblast growth factor 23-Klotho axis and bone turnover dynamics. Exp Physiol. 2025:10.1113/EP092401.
18. Vogt J, Föller M. Regulation of αKlotho. Cell Physiol Biochem. 2025;59:511-24.
19. Vázquez-Sánchez S, Blasco A, Polo-Salguero M, et al. Fibroblast growth factor 23 as a prognostic biomarker in post-myocardial infarction outcomes: influence of renal function and its modulation by Klotho. J Am Heart Assoc. 2026;15:e045175.
20. Liu M, Xia P, Tan Z, et al. Fibroblast growth factor-23 and the risk of cardiovascular diseases and mortality in the general population: a systematic review and dose-response meta-analysis. Front Cardiovasc Med. 2022;9:989574.
21. Paul S, Wong M, Akhabue E, et al. Fibroblast growth factor 23 and incident cardiovascular disease and mortality in middle-aged adults. J Am Heart Assoc. 2021;10:e020196.
22. Xue C, Yang B, Zhou C, et al. Fibroblast growth factor 23 predicts all-cause mortality in a dose-response fashion in pre-dialysis patients with chronic kidney disease. Am J Nephrol. 2017;45:149-59.
23. Musgrove J, Wolf M. Regulation and effects of FGF23 in chronic kidney disease. Annu Rev Physiol. 2020;82:365-90.
24. Gao S, Xu J, Zhang S, Jin J. Meta-analysis of the association between fibroblast growth factor 23 and mortality and cardiovascular events in hemodialysis patients. Blood Purif. 2019;47(Suppl 1):24-30.

Original Article

DOI: 10.4274/haseki.galenos.2026.63383
Med Bull Haseki 2026;64(4):229-236

Expert Evaluation of Large Language Models for Caregiver Counseling in Pediatric Seizures: A Comparative Study of ChatGPT, Gemini, and DeepSeek

Sema Bozkaya Yilmaz[1], Gunce Basarir[2]

[1]*University of Health Sciences Türkiye, Bursa City Hospital, Clinic of Pediatric Neurology, Bursa, Türkiye*

[2]*University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Pediatric Neurology, Istanbul, Türkiye*

Abstract

Aim: Caregiver-facing large language model (LLM) responses to pediatric seizure questions must provide accurate, complete, and safety-critical guidance because omissions may affect urgent decision-making. This study aimed to evaluate the clinical accuracy, completeness, readability, empathy, and reproducibility of responses generated by three LLMs to standardized caregiver questions about pediatric seizures.

Methods: This cross-sectional comparative study evaluated GPT-5.5, Gemini 3 Flash, and DeepSeek V3. Ten standardized caregiver questions addressing concerns about childhood seizures were submitted to each model in two runs during a 48-hour testing window in May 2026. Model identities were removed, and responses were evaluated by two pediatric neurologists using a 5-point rubric covering clinical accuracy and safety, completeness, readability and understandability, and empathy and tone. Inter-run consistency was assessed using Spearman's rank correlation coefficient.

Results: A total of 60 artificial intelligence-generated responses were evaluated. Expert-rated performance differed among the models in this exploratory analysis. Gemini received the highest overall scores within this prompt set and the May 2026 testing window, with a median score of 5.00; it was followed by GPT-5.5 and DeepSeek V3, which had median scores of 4.50 and 4.00, respectively. Although all models produced fluent and understandable responses, some outputs omitted clinically important safety information, including the five-minute emergency threshold. Inter-run consistency was moderate (Spearman's ρ=0.517). These findings are exploratory and may differ across model updates, testing periods, and clinical scenarios.

Conclusion: In this targeted expert evaluation of caregiver-oriented responses to pediatric seizures, model performance varied across LLMs, and repeated responses demonstrated only moderate consistency. Although Gemini achieved the highest scores within the prompt set and the testing window, the occasional omission of safety-critical advice, including the five-minute emergency threshold, indicates that LLMs should be used as supplementary sources of general information rather than as stand-alone tools for urgent seizure-related decisions. Relative model performance may change with model updates and when different clinical scenarios are evaluated.

Keywords: Pediatric seizures, large language models, ChatGPT, Gemini, DeepSeek, artificial intelligence, caregiver counseling, patient education

Introduction

Seizures are transient episodes of signs and/or symptoms resulting from abnormal excessive or synchronous neuronal activity in the brain (1). Febrile and non-febrile seizures are among the most common neurological emergencies in childhood. Because they often occur unexpectedly, they may cause considerable anxiety in parents, particularly after a first episode (2,3). Lacking immediate professional support, families are often driven by a sense of urgency to turn toward digital platforms for health-related guidance.

As large language model (LLM) applications become part of daily life, they are redefining how caregivers

Corresponding Author: Sema Bozkaya Yilmaz, MD, University of Health Sciences Türkiye, Bursa City Hospital, Clinic of Pediatric Neurology, Bursa, Türkiye
E-mail: semabozkayayilmaz@gmail.com **ORCID:** orcid.org/0000-0002-5389-5616
Received: 21.05.2026 **Accepted:** 06.08.2026 **Publication Date:** 24.09.2026

Cite this article as: Bozkaya Yilmaz S, Basarir G. Expert evaluation of large language models for caregiver counseling in pediatric seizures: a comparative study of ChatGPT, Gemini, and DeepSeek. Med Bull Haseki. 2026;64(4):229-236

navigate health-related information. While artificial intelligence (AI) technology has long been a staple in medicine, today's chat systems have become an increasingly accessible source of information for the general public. Recent studies and reviews have shown that LLMs are increasingly evaluated as tools for answering patient questions, generating patient education materials, and supporting caregiver-facing information; however, their performance remains variable, and concerns persist regarding accuracy, completeness, and safety (4-7). Built on distinct technical frameworks, ChatGPT, Gemini, and DeepSeek have emerged as some of the most accessible models for the general public today. Although LLMs have been increasingly studied in health care, their use in pediatric neurology remains less well characterized. Evidence is especially limited for caregiver-facing seizure counseling, where the quality of advice depends not only on accuracy but also on clear safety guidance (8).

Therefore, this study aimed to assess the clinical validity and communication quality of three widely used LLMs. Using standardized caregiver questions about pediatric seizures, we evaluated model-generated responses in terms of clinical accuracy, completeness, readability, empathy, and reproducibility. We hypothesized that the evaluated LLMs would differ in overall expert-rated performance and that repeated responses to identical caregiver prompts would show variability in clinically relevant content.

Materials and Methods

Compliance with Ethical Standards

This study did not involve human participants, animal subjects, patient-level data, private medical records, biological samples, identifiable personal data, patient images, or any clinical intervention. The prompts were standardized hypothetical caregiver questions developed by the authors, and the evaluated material consisted only of responses generated by publicly accessible LLMs. Therefore, institutional ethics committee approval and informed consent were not required. The authors declare no conflicts of interest and no financial support was received for this study.

Study Design and AI Models

This was a cross-sectional, prompt-based, comparative evaluation study designed to assess the quality and reproducibility of caregiver-oriented responses generated by LLMs in response to pediatric seizure-related questions. Three widely accessible LLMs were evaluated: ChatGPT/GPT-5.5 (OpenAI), Gemini 3 Flash (Google), and DeepSeek V3.

We used ten caregiver prompts, selected to represent common seizure-related questions in pediatric neurology practice. These prompts addressed acute seizure first aid, febrile seizures, emergency referral, electroencephalography interpretation, antiseizure treatment, neurodevelopmental concerns, the risk of later epilepsy, and daily-life precautions. Each question was entered twice into each of the three models, yielding 60 responses in total. The prompts were submitted through the publicly available web interfaces of the models without changing the default settings. To limit the effects of possible model updates, all queries were completed within a single 48-hour period in May 2026. Before the second run, previous conversations were cleared so that earlier prompts would not influence later responses.

Model names and versions were recorded as displayed in the public web interfaces at the time of testing. Because LLMs are frequently updated, the findings reflect model behavior during this specific testing period and may not be directly reproducible at later dates. The study workflow included prompt development, two independent model runs, collection and de-identification of responses, blinded expert evaluation, and statistical analysis, as summarized in Figure 1.

Prompt Development and Reference Standards

The prompts were developed by two pediatric neurology specialists to reflect common caregiver questions encountered in clinical practice. Rather than using a large number of automated prompts, the study deliberately focused on ten high-priority questions that would allow detailed expert assessment of clinically relevant and potentially safety-sensitive content. The full list of prompts is presented in Table 1.

Before formal scoring, the authors defined core expected content for seizure-related caregiver counseling based on established seizure terminology, febrile seizure guidance, and commonly accepted seizure first-aid principles. These predefined reference standards covered seizure first aid, emergency warning signs, febrile seizure counseling, diagnosis and treatment, developmental and epilepsy risks, and daily-life precautions. They were used to support the clinical accuracy, safety, and completeness domains of the scoring rubric. The predefined reference standards are summarized in Table 2.

Outcomes

The primary outcome was the overall expert-rated performance score assigned to each LLM-generated response. Secondary outcomes included domain-specific scores for clinical accuracy and safety; completeness; readability and understanding; and empathy and tone. Additional secondary outcomes included inter-run consistency of repeated model outputs, inter-rater agreement between the two evaluators, and presence of safety-critical omissions, including failure to mention the five-minute emergency threshold when clinically relevant.

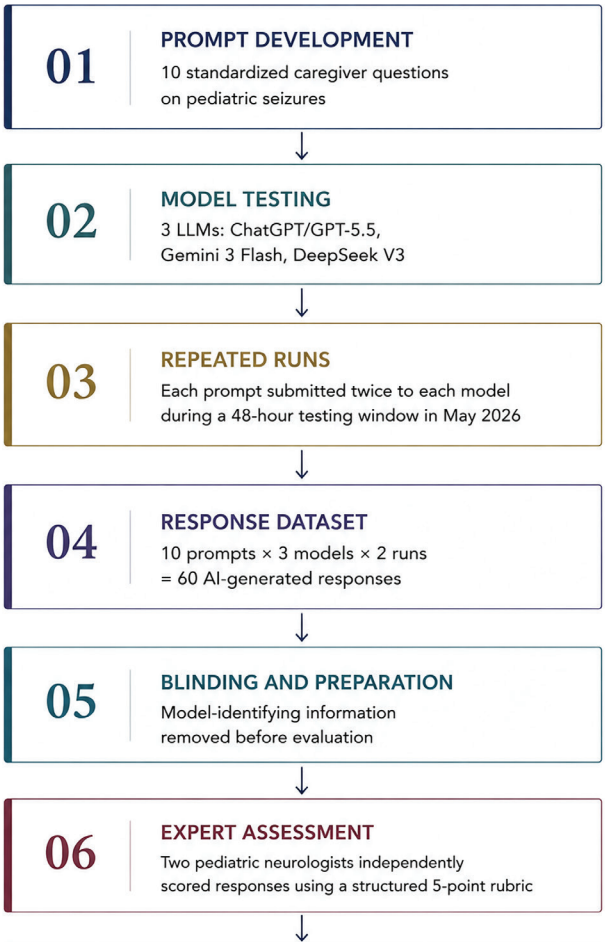

Figure 1. Study workflow. The workflow summarizes prompt development, model testing, repeated response generation, response de-identification, blinded expert evaluation, and statistical analysis. Ten standardized caregiver prompts were submitted twice to each of three large language models, resulting in 60 AI-generated responses
AI: Artificial intelligence

Evaluation Protocol

The responses generated by each model were collected in a standardized document. All information that could identify the source model was removed before evaluation to ensure that the assessors were blinded to model identity. However, complete blinding could not be guaranteed because model outputs may have recognizable stylistic characteristics.

The responses were independently evaluated by two pediatric neurologists. Before formal scoring, a pilot evaluation was performed using sample responses to align evaluators and ensure consistent interpretation of scoring domains. Each response was scored using a 5-point rubric ranging from 1, indicating very poor performance, to 5, indicating excellent performance. The rubric covered four domains: clinical accuracy and safety, completeness, readability and understanding, and empathy and tone. The scoring rubric is presented in Table 3. Rubric development followed a prespecified, stepwise process: the two pediatric neurologists first identified the clinically and communicatively essential domains; mapped the accuracy/safety and completeness domains to the predefined seizure reference standards; reviewed the readability and empathy domains in relation to dimensions commonly assessed in prior expert evaluations of patient- and caregiver-facing LLM outputs (5-7); and then pilot-tested the draft rubric on sample responses to harmonize domain interpretation and scoring thresholds before formal evaluation.

Each evaluator assigned four domain scores and a separate holistic overall performance score (1-5) for each response. The two evaluators' overall performance scores were averaged to yield the primary overall score for that response. Scores within each domain were likewise averaged across the two evaluators and reviewed descriptively. This overall score was used for descriptive model comparisons. The domain-specific scores were also reviewed to identify clinically relevant strengths and weaknesses in the generated responses.

Although the rubric was not a formally validated instrument, it was developed to reflect clinically relevant dimensions of caregiver-facing medical information, including clinical accuracy and safety, completeness, readability, and empathic communication. The four domains were defined a priori and selected because they correspond to recurring assessment dimensions in published evaluations of LLM-generated health information (5-7). The clinical accuracy, safety, and completeness domains were anchored to the predefined seizure-related reference standards described above, whereas readability and empathy were assessed as caregiver-oriented communication domains. Following the pilot review, the evaluators discussed discrepancies in domain interpretation and agreed on the operational meaning of each score level; the final scoring criteria and anchors are provided in Table 3. Therefore, the scoring process combined structured expert assessment with clinically anchored reference criteria.

Statistical Analysis

All data were analyzed using jamovi version 2.3 (The jamovi project, Sydney, Australia) and JASP version 0.18.3 (JASP Team, Amsterdam, the Netherlands). Descriptive statistics were reported as medians and interquartile ranges (IQR) because the score distributions were nonparametric and exhibited restricted ranges in some domains.

Table 1. Standardized caregiver prompts used for LLM evaluation

Prompt no.	Clinical topic	Standardized caregiver question
1	Acute seizure management	"My child had a seizure. What should I do?"
2	Febrile seizures-prognosis	"Are febrile seizures harmful?"
3	Febrile seizures-recurrence	"Can febrile seizures recur?"
4	Antiseizure medication	"Does a child who has had a seizure need medication?"
5	EEG interpretation	"Can seizures still occur if the EEG is normal?"
6	Emergency evaluation	"When should I take my child to the hospital?"
7	Seizure recognition	"How can I recognize a seizure in my child? What are the signs of seizures?"
8	Neurodevelopmental outcome	"Can seizures affect my child's intelligence or development?"
9	Epilepsy risk after febrile seizures	"Can febrile seizures develop into epilepsy later in life?"
10	Daily life precautions	"What precautions should be taken in the daily life of a child with seizures (sports, bathing, sleep)?"

EEG: Electroencephalography

Table 2. Predefined reference standards for seizure-related caregiver counseling

Domain	Expected core elements
Seizure first aid	Safe positioning, protection from injury, no restraint, nothing placed in the mouth, observation of breathing and seizure duration
Emergency warning signs	Urgent medical evaluation for seizures lasting approximately ≥5 minutes, recurrent seizures without recovery, breathing difficulty/cyanosis, prolonged altered consciousness, focal deficit, or first seizure
Febrile seizure counseling	Simple febrile seizures are usually benign; recurrence is possible; risk assessment depends on age, seizure features, neurological status, and family history
Diagnosis and treatment context	A normal EEG does not exclude seizures; antiseizure medication decisions depend on seizure type, recurrence risk, etiology, EEG findings, and specialist evaluation
Developmental and epilepsy risk	Long-term outcome depends on seizure type, underlying cause, seizure burden, and associated neurodevelopmental or neurological findings
Daily-life precautions	Supervision during bathing/swimming, avoidance of unsafe heights or high-risk activities, attention to sleep and medication adherence when relevant, and encouragement of age-appropriate activities

The reference standards were defined before formal scoring and were based on established seizure terminology, febrile seizure guidance, and commonly accepted seizure first-aid principles
EEG: Electroencephalography

Table 3. Scoring rubric for evaluating LLM-generated responses

Score	Clinical accuracy & safety	Completeness (adequacy)	Readability & understanding	Empathy & tone
5 (excellent)	Perfectly aligned with ILAE/AAP guidelines. No medical errors. Correct dosages/protocols.	All critical aspects (red flags, 5-min rule) covered comprehensively.	Professional yet perfectly accessible for laypersons. No jargon.	Highly supportive, reassuring, and culturally sensitive tone.
4 (good)	Accurate and safe advice. Only minor, non-critical details missing.	Covers all main points but lacks some secondary depth.	Clear and easy to read. Very few technical terms.	Reassuring and appropriate for a medical context.
3 (fair)	Generally safe but lacks specific clinical depth. Minor omissions in guidance.	Missing important secondary information (e.g., specific lifestyle tips).	Understandable but uses some medical jargon that may confuse parents.	Neutral tone. Helpful but lacks proactive empathy.
2 (poor)	Contains omissions that could lead to confusion or delayed action.	Significant gaps. Fails to mention several key clinical points.	Difficult to follow. Overly technical or excessively brief.	Clinical or dismissive tone. Fails to address parental anxiety.
1 (very poor)	Clinically unsafe. Incorrect dosages or dangerous advice.	Fails to answer the core question. Minimal or irrelevant content.	Unintelligible or completely inappropriate for a caregiver.	Apathetic, alarming, or inappropriate tone.

Scores ranged from 1 to 5, with higher scores indicating better performance
AAP: American Academy of Pediatrics, ILAE: International League Against Epilepsy, LLM: large language model

The Kruskal-Wallis H test was used for exploratory comparison of overall performance scores among the three models, with post-hoc comparisons performed using the Dwass-Steel-Critchlow-Fligner (DSCF) procedure. Because the same standardized prompts were repeated across models and runs, the observations were not fully independent. A mixed-effects model or a generalized estimating equations approach would ordinarily be preferable for this repeated-prompt structure. However, these models were not fitted because the dataset included only ten prompt clusters, two runs per model, and strongly restricted score distributions with a ceiling effect, conditions that could yield unstable or poorly identified estimates. Therefore, statistical comparisons were interpreted as exploratory and considered together with descriptive score distributions and clinically relevant response patterns rather than as confirmatory evidence. Accordingly, the reported p-values are only hypothesis-generating and should not be interpreted as confirmatory evidence of superiority between models.

Inter-run consistency was assessed by correlating the mean overall scores from Run 1 and Run 2 across 30 matched prompt-model pairs using Spearman's rank correlation coefficient (ρ); exact agreement was also reported descriptively. Inter-rater agreement was primarily assessed using absolute percentage agreement. Because score distributions were restricted and skewed in some qualitative domains, particularly empathy and completeness, chance-corrected agreement statistics were difficult to interpret. Therefore, absolute percent agreement was reported as a descriptive measure of inter-rater agreement and was not considered a substitute for chance-corrected reliability statistics.

To further describe clinical consensus, scores were also interpreted using a dichotomous framework: scores of 3-5 were considered clinically acceptable, whereas scores of 1-2 were considered unacceptable. A p-value below 0.05 was used as a threshold for statistical significance, but all inferential findings were interpreted cautiously in light of the exploratory design and repeated-prompt structure. No causal, confirmatory, or broadly generalizable inferences were made from these analyses.

Results

Comparative Performance of AI Models

A total of 60 AI-generated responses were evaluated, with 20 responses obtained from each model. These responses were produced by submitting ten standardized prompts to each model in two independent runs. In the exploratory comparison, overall performance scores differed among the three models (Kruskal-Wallis $\chi^2(2)=54.6$, $p<0.001$). Gemini had the most favorable overall ratings in this prompt set, with scores clustering at the maximum value (median 5.00, IQR: 5.00-5.00). GPT-5.5 had a median of 4.50 (IQR: 4.50-4.50), whereas DeepSeek V3 had lower overall scores, with a median of 4.00 (IQR: 4.00-4.00). The distribution of overall performance scores across the three models is shown in Figure 2. Within this prompt set, Gemini received the maximum score for all evaluated responses, indicating a ceiling effect in overall performance scores. This pattern describes only the tested prompts and the May 2026 model versions and should not be interpreted as stable superiority across updates or other clinical contexts.

Post-hoc Pairwise Comparisons

In exploratory post-hoc comparisons using the DSCF procedure, Gemini received higher overall performance scores than GPT-5.5 ($W=8.73$, $p<0.001$) and DeepSeek V3 ($W=8.73$, $p<0.001$), while GPT-5.5 received higher scores than DeepSeek V3 ($W=7.44$, $p<0.001$). Qualitatively, GPT-5.5 responses were generally easy for caregivers to read and understand, whereas some responses contained less clinical detail than Gemini responses; these domain-level observations were descriptive and were not tested statistically. Which focused on emergency evaluation: GPT-5.5 used an authoritative tone but did not explicitly mention the five-minute threshold for emergency evaluation; this omission is clinically relevant because prolonged seizures require urgent medical assessment.

Inter-rater Reliability Analysis

Overall, a high level of clinical consensus was observed between the two blinded evaluators. The restricted score range-particularly because one evaluator frequently

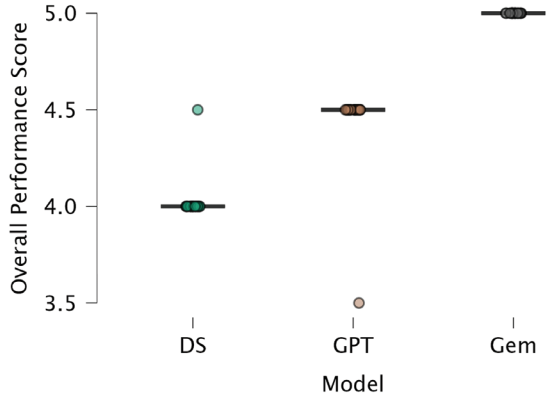

Figure 2. Distribution of overall performance scores across AI models. The plot illustrates the distribution of expert-rated overall performance scores for DeepSeek, GPT-5.5, and Gemini. Individual data points and boxplots are shown for each model. Gemini scores clustered at the maximum value, indicating a ceiling effect within the present prompt set
AI: Artificial intelligence

assigned maximum scores in some domains-limited the usefulness of chance-corrected agreement statistics. Because of the restricted score range and lack of variance in these categories, a formal Cohen's kappa coefficient could not be meaningfully calculated. Therefore, absolute percent agreement was reported descriptively. In the completeness domain, the absolute percent agreement was 66.7%. In the empathy domain, evaluators appeared to apply different subjective thresholds, with scores of 3-5 versus 4-5; the exact-match agreement was 33.3%. Although exact agreement for empathy was lower, the discrepancy reflected differences in subjective communication thresholds rather than a major disagreement regarding clinical safety. When scores were dichotomized into clinically acceptable and unacceptable categories, inter-rater agreement was 100%. This suggests that the lower exact-match rates mainly reflect differences in scoring thresholds for communication quality, rather than disagreement about clinical safety. Despite variability in exact scores, no response from any model was rated as clinically poor (1) or unsafe (2) by either evaluator in any evaluated category.

Reliability and Reproducibility Analysis

Across the 30 matched prompt-model pairs, mean overall scores were strongly correlated between the first and second runs (Spearman's ρ=0.922, $p<0.001$), with exact agreement in 28 pairs (93.3%). Exact agreement was observed in 9/10 GPT-5.5 pairs, 10/10 Gemini pairs, and 9/10 DeepSeek V3 pairs. Nevertheless, repeated responses were not textually identical, and qualitative differences in detail and emphasis were observed across runs. This variability is relevant because caregivers may receive different safety information depending on when and how the same question is asked. All generated responses were successfully recorded and included in the analysis.

Discussion

Seizures are among the most frequent neurological emergencies in children, representing nearly 3% of pediatric emergency visits and affecting about 5% of children worldwide (9,10). For many parents, witnessing a seizure is intensely frightening, and some describe it as a moment in which they fear losing their child (11,12). In moments of uncertainty, families may seek quick, accessible guidance before they can reach a clinician. With conversational AI tools now widely available, caregivers may also turn to LLM-based systems for explanations, reassurance, or practical advice. For this reason, incomplete or inaccurate AI-generated advice in pediatric seizure counseling requires careful evaluation.

In the present exploratory evaluation, overall expert-rated performance differed among the evaluated AI models. For this specific set of prompts, Gemini received higher overall performance scores than GPT-5.5 and DeepSeek V3. The variability in performance across the evaluated models is consistent with recent systematic reviews showing that LLMs may provide useful patient-facing medical information but remain limited by inconsistent performance, incomplete responses, accuracy gaps, and safety concerns. Qualitatively, Gemini's responses more consistently included guideline-concordant elements, including International League Against Epilepsy-based seizure terminology and American Academy of Pediatrics-aligned recommendations for febrile seizure management; however, this was a descriptive observation and was not tested statistically. In addition, the overall performance result should be interpreted within the limitations of the selected prompt set and the testing window. The ceiling effect observed with Gemini's uniform maximum scores may also indicate that the 5-point rubric was not sensitive enough to distinguish between strong responses at the upper end of performance. Thus, Gemini's higher ratings should be understood as a context-specific finding from this prompt set and testing period rather than evidence of general or enduring model superiority; rankings may change after model updates or when different pediatric seizure scenarios are tested.

Although all three models were able to generate fluent and understandable responses, our findings show that linguistic fluency should not be assumed to reflect clinical reliability. One notable observation was that responses could sound fluent and authoritative while still missing clinically relevant content-a pattern we describe as a potential "confidence gap". For example, GPT-5.5's response to Prompt 6 was clearly written and generally helpful but did not explicitly mention the five-minute threshold for emergency evaluation. This omission is clinically relevant because seizures lasting five minutes or longer are generally treated as prolonged seizures and require urgent medical evaluation. Although no response was rated as manifestly unsafe, such omissions may nevertheless be important in seizure counseling for caregivers, where urgent warning signs must be recognized promptly. For caregivers without medical training, a confidently written answer may appear trustworthy even when it is incomplete.

The reproducibility analysis added another important dimension to clinical reliability. The pooled Spearman correlation of 0.922 indicated strong agreement between run-level overall scores. However, because score ranges were highly restricted within models and the pooled correlation was partly influenced by between-model differences, this finding should not be interpreted as

complete content-level reproducibility. Repeated responses were not textually identical and differed in detail and emphasis; such variation may affect the communication of safety warnings, emergency thresholds, and follow-up recommendations.

Differences in scoring for subjective domains, such as empathy, also highlight the difficulty of standardizing the quality of communication. What one evaluator considers sufficiently reassuring may be judged as less supportive by another. Similar concerns were reported in caregiver-oriented LLM evaluations, where models produced clear and accessible responses but still omitted relevant practical details when compared with professional reference standards (6). Future studies should include caregivers directly, since expert-rated quality may not fully reflect how families understand, trust, or act upon AI-generated advice in real-life settings. It will also be important to determine whether these responses reduce anxiety or instead increase confusion, false reassurance, or unnecessary healthcare-seeking behavior.

Study Limitations

Several limitations should be acknowledged. The study was based on ten standardized clinical scenarios rather than real-life caregiver interactions. This focused prompt set enabled expert neurologists to examine high-stakes, seizure-related questions in detail; however, the findings should not be generalized to all pediatric seizure-related situations. Specifically, the study was limited to ten standardized prompts, two expert evaluators, and a single 48-hour testing window in May 2026. Consequently, the results may not represent the full range of caregiver wording, seizure types, levels of clinical urgency, sociocultural contexts, or longitudinal changes in model behavior. Because LLMs are updated frequently, these findings should be considered a snapshot of model performance during the specific testing period and may not be generalizable to future model versions. In addition, because the dataset was based on repeated standardized prompts, the observations were not fully independent; therefore, the statistical comparisons should be interpreted as exploratory rather than confirmatory.

Although explicit model identifiers were removed prior to evaluation, complete blinding could not be guaranteed, as some models may exhibit recognizable stylistic response patterns. The responses were evaluated by expert pediatric neurologists using a structured rubric; however, the rubric was not a formally validated instrument. In addition, the study did not directly assess caregivers' perceptions, comprehension, trust, anxiety, healthcare-seeking behavior, or real-life decision-making after receiving AI-generated advice. Therefore, expert-rated quality may not fully reflect how caregivers would understand or act upon these responses during an actual seizure-related situation. Finally, the study evaluated text-based responses only and did not examine voice-based, multimodal, or interactive use of LLMs in real-world settings. Despite these limitations, the study has several strengths, including the use of standardized clinically relevant prompts, predefined reference standards, blinded independent evaluation by two pediatric neurologists, and repeated model testing to assess reproducibility. These methodological features provide a structured and clinically grounded comparison of LLM performance in caregiver-oriented pediatric seizure counseling.

Conclusion

In this targeted expert evaluation of 60 LLM-generated responses to caregiver questions about pediatric seizures, Gemini achieved the highest scores within the tested prompt set. However, overall scores were highly consistent across runs, although repeated responses differed in detail and emphasis, and some responses omitted clinically important emergency advice, including the five-minute emergency threshold. This comparative finding is limited to the May 2026 testing window and to the clinical scenarios examined and may change with model updates or alternative prompt sets. These findings suggest that LLMs may support general caregiver education, but they should not be used as stand-alone sources for urgent seizure-related decisions. Future studies should examine how caregivers understand, trust, and act upon AI-generated seizure advice in real-life settings.

Ethics

Ethics Committee Approval: This study did not involve human participants, animal subjects, or the use of private medical records. As the evaluation was conducted using publicly available large language models and standardized clinical prompts, institutional review board approval was not required.

Informed Consent: Informed consent was not required, as the study did not involve human participants or the collection or use of identifiable patient-level data.

Footnotes

Authorship Contributions

Concept: S.B.Y., G.B., Design: S.B.Y., G.B., Data Collection or Processing: S.B.Y., G.B., Analysis or Interpretation: S.B.Y., G.B., Literature Search: S.B.Y., G.B., Writing: S.B.Y., G.B.

Conflict of Interest: The authors declared that there were no conflicts of interest.

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

Declaration Regarding the Use of AI and AI-Assisted Technologies: This study evaluates responses

generated by GPT-5.5, Gemini 3 Flash, and DeepSeek V3. The study design, prompt selection, scoring framework, clinical evaluation, statistical analysis, interpretation of the findings, and final manuscript decisions were carried out by the authors. AI tools did not replace author judgment at any stage. The authors reviewed and approved the final manuscript and take full responsibility for its content.

References

1. Fisher RS, van Emde Boas W, Blume W, et al. Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). Epilepsia. 2005;46:470-2.
2. Wirrell E, Turner T. Parental anxiety and family disruption following a first febrile seizure in childhood. Paediatr Child Health. 2001;6:139-43.
3. Save-Pédebos J, Bellavoine V, Goujon Eet al. Difference in anxiety symptoms between children and their parents facing a first seizure or epilepsy. Epilepsy Behav. 2014;31:97-101.
4. Yun HS, Bickmore T. Online health information-seeking in the era of large language models: cross-sectional web-based survey study. J Med Internet Res. 2025;27:e68560.
5. Jacobs MMG, Oosterhoff JHF, Agricola R, van der Weegen W. Large language models versus healthcare professionals in providing medical information to patient questions: a systematic review. Int J Med Inform. 2026;209:106250.
6. Pérez-Esteve C, Guilabert M, Matarredona V, et al. AI in home care-evaluation of large language models for future training of informal caregivers: observational comparative case study. J Med Internet Res. 2025;27:e70703.
7. Khosravi M, Zamaninasab Z, Mojtabaeian SM, Dindar Demiray EK, Arab-Zozani M. A systematic review of the limitations of large language models in generating healthcare content. PLOS Digit Health. 2026;5:e0001354.
8. Fennig U, Yom-Tov E, Savitzky L, et al. Bridging the conversational gap in epilepsy: using large language models to reveal insights into patient behavior and concerns from online discussions. Epilepsia. 2025;66:686-99.
9. Smith RA, Martland T, Lowry MF. Children with seizures presenting to accident and emergency. Emerg Med J. 1996;13:54-8.
10. Subcommittee on Febrile Seizures; American Academy of Pediatrics. Neurodiagnostic evaluation of the child with a simple febrile seizure. Pediatrics. 2011;127:389-94.
11. Hall-Parkinson D, Tapper J, Melbourne-Chambers R. Parent and caregiver knowledge, beliefs, and responses to convulsive seizures in children in Kingston, Jamaica: a hospital-based survey. Epilepsy Behav. 2015;51:306-11
12. Besag FM, Nomayo A, Pool F. The reactions of parents who think that a child is dying in a seizure–in their own words. Epilepsy Behav. 2005;7:517-23.

Original Article

DOI: 10.4274/haseki.galenos.2026.85547
Med Bull Haseki 2026;64(4):237-243

Long-term Outcomes According to Final Single-stent or Two-stent Treatment for Culprit Medina 1.1.1 Bifurcation Lesions in ST-elevation Myocardial Infarction

Fahrettin Turna, Direnc Yilmaz

Sakarya University Training and Research Hospital, Department of Cardiology, Sakarya, Türkiye

Abstract

Aim: Long-term data comparing stenting strategies for culprit bifurcation lesions during primary percutaneous coronary intervention (PCI) are limited. We compared final single- and two-stent strategies in patients with ST-elevation myocardial infarction (STEMI) and Medina 1.1.1 lesions.

Methods: We retrospectively reviewed 2,452 patients undergoing primary PCI between January 2018 and December 2019. The study included 112 patients: 46 had a final single-stent result, and 66 were treated with two stents. The primary endpoint was the first occurrence of target lesion revascularization (TLR) or all-cause death. Kaplan-Meier, log-rank test, Cox regression, and logistic regression analyses were used.

Results: Mean follow-up was 61.5±22.4 months. The primary endpoint occurred in 12 patients in the single-stent group and 13 patients in the two-stent group. Event-free survival did not differ significantly [log-rank p=0.307; hazard ratio, 0.67; 95% confidence interval (CI), 0.30-1.47]. Each 5% decrease in ejection fraction was associated with higher odds of the composite endpoint (adjusted odds ratio, 1.31; 95% CI, 1.05-1.63; p=0.017).

Conclusion: Long-term outcomes were not significantly different between the two strategies. Lower ejection fraction remained associated with the composite endpoint after adjustment. Strategy selection may be guided by lesion anatomy and the procedural result.

Keywords: ST-elevation myocardial infarction, percutaneous coronary intervention, coronary bifurcation, provisional stenting, two-stent strategy

Introduction

Primary percutaneous coronary intervention (PCI) is the preferred reperfusion strategy for ST-elevation myocardial infarction (STEMI). When the culprit lesion involves a coronary bifurcation, treatment is more challenging because rapid restoration of main-vessel flow must be balanced against preservation of a clinically relevant side branch (1). The Medina classification defines bifurcation involvement according to disease in the proximal main vessel, distal main vessel, and side branch (2). During STEMI, thrombus burden, impaired baseline flow, vasoconstriction, bifurcation geometry, and hemodynamic instability may further complicate procedural decision-making (3-5).

A provisional single-stent approach is generally preferred for many bifurcation lesions, whereas planned two-stent treatment may be appropriate in selected complex anatomies. Long-term findings from EBC MAIN supported a stepwise provisional strategy in many true left-main bifurcations, while pooled DKCRUSH X data

Corresponding Author: Fahrettin Turna, MD, Sakarya University Training and Research Hospital, Department of Cardiology, Sakarya, Türkiye
E-mail: fturna_53@hotmail.com **ORCID:** orcid.org/0000-0003-3483-9456
Received: 29.04.2026 **Accepted:** 01.08.2026 **Publication Date:** 24.09.2026

Cite this article as: Turna F, Yilmaz D. Long-term outcomes according to final single-stent or two-stent treatment for culprit medina 1.1.1 bifurcation lesions in ST-elevation myocardial infarction. Med Bull Haseki. 2026;64(4):237-243

favored systematic DK-crush in selected complex lesions (6,7). The STEMI-BIF registry added a different perspective by highlighting the potential importance of active side-branch protection during primary PCI (8). More recent studies suggested that intravascular imaging and active side-branch protection may improve procedural outcomes, although evidence specific to STEMI remains limited (9,10).

We hypothesized that long-term outcomes would not differ significantly between patients with a final single-stent result and those treated with two stents. We, therefore, compared the time to first target-lesion revascularization or all-cause death between the two strategies in patients with culprit Medina 1.1.1 bifurcation lesions undergoing primary PCI. We also examined clinical factors associated with the composite endpoint. This study may contribute to clinical decision-making by providing long-term real-world data supporting an anatomy-guided and procedural-result-guided choice between final single-stent and two-stent strategies in this high-risk setting.

Materials and Methods

Compliance with Ethical Standards

The study protocol was approved by the Sakarya University Local Institutional Ethics Committee (approval no: 71522473/050.01.04/617, date: 20.11.2020). The study was conducted in accordance with the Declaration of Helsinki. Because the study was retrospective and all patient data were analyzed anonymously, the requirement for written informed consent was waived by the ethics committee.

Study Design and Patient Selection

We retrospectively reviewed patients who underwent primary PCI for STEMI at our center between January 2018 and December 2019. Angiography records, procedural reports, hospital charts, electronic medical records, and follow-up files were examined. The patient selection process is shown in Figure 1.

Patients were included if they had undergone primary PCI for an index STEMI and had a true bifurcation lesion classified as Medina 1.1.1. The side branch was considered clinically relevant when its reference diameter was greater than 2.5 mm, or when it exceeded 2.0 mm and supplied a substantial myocardial territory. All included patients had been treated with drug-eluting stents and had complete clinical, angiographic, procedural, and follow-up data.

Patients were excluded if the treated bifurcation had a Medina classification other than 1.1.1, if the bifurcation lesion was not responsible for the index STEMI, or if angiographic documentation was insufficient to confirm lesion anatomy. Patients with missing procedural or follow-up data were also excluded. In addition, surviving patients with documented absence or premature discontinuation of dual antiplatelet therapy during the first year were not included.

Angiographic Assessment and Percutaneous Coronary Intervention Strategy

All coronary angiograms were reviewed to confirm the culprit bifurcation lesion and its Medina 1.1.1 anatomy. The following procedural variables were recorded: bifurcation angle, main-vessel stent diameter and length, presence of multivessel disease, final kissing balloon inflation, side-branch loss, and conversion from an initially provisional approach to a two-stent technique.

The choice of stenting strategy was left to the treating operator. Decisions were based on side-branch size, the amount of myocardium supplied by the side branch, lesion morphology, thrombus burden, baseline coronary flow, bifurcation angle, procedural feasibility, and the patient's clinical condition.

In patients treated through a provisional pathway, the main vessel was stented first. Side-branch balloon dilatation, proximal optimization technique (POT), POT-side-POT, final kissing balloon inflation, or implantation of a second stent were performed when considered necessary according to the angiographic results. Two-stent techniques include mini-crush, T-and-protrusion (TAP), and culotte.

Patients who initially underwent provisional treatment but subsequently required a second stent were assigned to the two-stent group based on the final procedural result. Guideline-directed medical treatment was prescribed after PCI. Dual antiplatelet therapy was planned to continue for at least 12 months, unless death or a clinically relevant contraindication occurred.

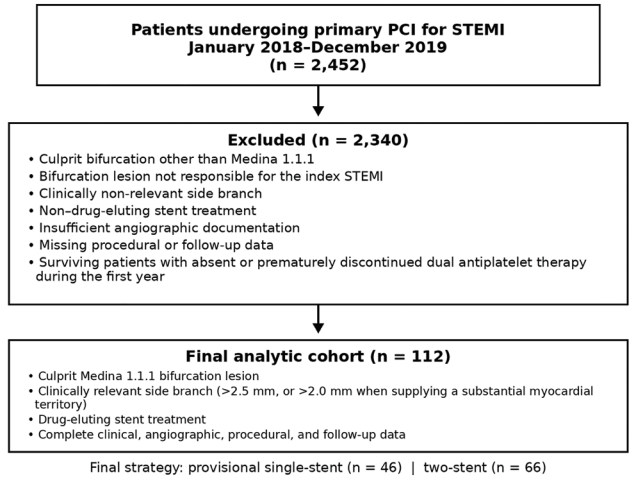

Figure 1. Patient flow diagram
PCI: Percutaneous coronary intervention, STEMI: ST-elevation myocardial infarction

Follow-up and Study Endpoints

Follow-up information was obtained from hospital records, outpatient files, electronic medical records, and telephone contact when necessary. Follow-up duration was recorded in months. For patients who died during follow-up, the follow-up period ended in the month of their death.

Target lesion revascularization (TLR) was defined as the repeat revascularization of the previously treated target lesion. The primary endpoint was the time from the index PCI to the first occurrence of TLR or all-cause death. Cause-specific mortality was not consistently available in the retrospective records; therefore, cardiovascular and non-cardiovascular deaths could not be analyzed separately.

When a patient experienced more than one qualifying event, only the first event was included in the Kaplan-Meier analysis. Subsequent events were not counted again. Patients without an event were censored at their last documented follow-up. Secondary time-to-event endpoints were all-cause death and first TLR, analyzed separately.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics for Windows, version 30.0 (IBM Corp., Armonk, NY, USA). Continuous variables were assessed for normality using visual methods (histograms and Q-Q plots) and the Shapiro-Wilk test. Normally distributed continuous variables were expressed as mean ± standard deviation and compared using Student's t-test, whereas non-normally distributed variables were presented as median (interquartile range) and compared using the Mann-Whitney U test. Categorical variables were reported as numbers and percentages and were compared using the chi-square test or Fisher's exact test, as appropriate.

Event-free survival was estimated using the Kaplan-Meier method, and the groups were compared with the log-rank test. A univariable Cox proportional hazards model was used to calculate the unadjusted hazard ratio for the primary endpoint in the final two-stent group compared with the final single-stent group. The proportional hazards assumption was assessed using Schoenfeld residuals, with no evidence of violation.

Exploratory univariable logistic regression analyses were performed to examine the associations of final stenting strategy, age, ejection fraction (EF), and diabetes mellitus with the composite endpoint. Variables associated with the endpoint in univariable analyses were considered for inclusion in the multivariable model. The final stenting strategy was retained regardless of its univariable p-value, because it was the main exposure of interest. Because only 25 patients reached the primary endpoint, the multivariable model was limited to final stenting strategy, EF, and diabetes mellitus. Ejection fraction was entered into the model per 5% decrease.

The results of the logistic regression analyses were reported as odds ratios with 95% confidence intervals (CIs). Model discrimination was assessed using the area under the receiver operating characteristic curve. Calibration was evaluated with the Hosmer-Lemeshow goodness-of-fit test using 10 groups based on deciles of predicted probability. The Akaike information criterion and McFadden's pseudo-R^2 were also calculated. All tests were two-sided, and a p-value below 0.05 was considered statistically significant.

Results

Study Population and Procedural Characteristics

Among 2,452 patients screened, 112 met the prespecified criteria and were included in the final analysis. The patient selection pathway is shown in Figure 1.

The final single-stent group included 46 patients, and the final two-stent group included 66 patients. Baseline characteristics were broadly comparable. Observed follow-up was 56.9±24.1 months in the single-stent group and 64.7±20.6 months in the two-stent group (p=0.070). Overall follow-up was 61.5±22.4 months, with a median of 67.5 months. Final kissing balloon inflation was more frequent in the two-stent group, whereas side-branch loss occurred only in cases with a final single-stent result. Baseline, procedural, and clinical characteristics were presented in Table 1.

In the two-stent group, the mini-crush technique was used in 30 patients, and TAP and culotte were each used in 18 patients. Among the 54 patients initially treated through a provisional pathway, 8 required a bailout conversion to a two-stent technique. The final procedural pathways among initially provisional cases were summarized in Table 2.

Time-to-event Outcomes

The primary endpoint occurred in 12 patients in the single-stent group and 13 in the two-stent group. Kaplan-Meier analysis showed no statistically significant difference in event-free survival (log-rank p=0.307; Figure 2). The unadjusted Cox hazard ratio for the two-stent strategy was 0.67 (95% CI, 0.30-1.47; p=0.316). Death-only and first-TLR analyses also did not differ significantly (log-rank p=0.393 and p=0.329, respectively).

Patients who reached the composite endpoint had lower baseline EF and a higher frequency of diabetes mellitus. Clinical characteristics according to composite endpoint status are shown in Table 3.

Regression Analyses

Lower EF and diabetes mellitus were associated with the composite endpoint in univariable logistic regression analysis (Table 4).

In the parsimonious multivariable model, EF remained significant, whereas final strategy and diabetes mellitus were not significant. Model discrimination was modest (area under the curve, 0.688). The Hosmer-Lemeshow test did not provide evidence of poor model fit (chi-square =4.13, df=8; p=0.845). Akaike information criterion was 116.04, and McFadden's pseudo-R^2 was 0.092. Multivariable estimates were presented in Table 5.

The adjusted estimates from the multivariable model were illustrated in Figure 3.

Discussion

In this cohort of patients with STEMI and a culprit Medina 1.1.1 bifurcation lesion, long-term clinical outcomes did not differ significantly between patients with a final single-stent result and those treated with two stents. The same pattern was observed in the analyses of the composite endpoint, all-cause death, and first TLR. These findings should not be taken as evidence that the two approaches are equivalent, because treatment was not randomized and the number of events was limited. They do, however, suggest that the final number of stents was less closely related to the outcome than the patient's underlying clinical risk. Among the variables examined, lower baseline EF was most consistently associated with the composite endpoint.

Bifurcation PCI during STEMI differs from elective procedures in several practical respects. The operator must restore flow quickly, often in the presence of thrombus, vasoconstriction, unstable hemodynamics, and poor visualization of the side-branch ostium. Under these conditions, a technique that is straightforward in an elective case may be difficult to complete or prolong the procedure unnecessarily. European Bifurcation Club recommendations support beginning with a provisional approach in many lesions and adding a second stent when side-branch flow, residual stenosis, or dissection remains unacceptable (3-5). Our results are consistent with this approach. Most patients who entered the provisional pathway were treated without a second stent, whereas eight required bailout conversion because the final angiographic result was unsatisfactory.

Table 1. Baseline, procedural, and clinical characteristics according to final stenting strategy

Parameter	Single-stent (n=46)	Two-stent (n=66)	p-value
Age, years	63.2±11.2	61.3±12.4	0.400
Follow-up duration, months	56.9±24.1	64.7±20.6	0.070
Main-vessel stent length, mm	21.3±8.1	23.4±6.6	0.139
Main-vessel stent diameter, mm	3.00±0.50	2.97±0.40	0.703
Bifurcation angle, degrees	54.2±16.5	53.1±14.9	0.721
EF, %	48.3±9.5	47.5±11.1	0.668
Male sex	33 (71.7)	53 (80.3)	0.290
Diabetes mellitus	18 (39.1)	20 (30.3)	0.331
Hypertension	31 (67.4)	49 (74.2)	0.429
Chronic kidney disease	8 (17.4)	9 (13.6)	0.603
Previous coronary artery disease	18 (39.1)	25 (37.9)	0.893
Multivessel disease	23 (50.0)	40 (60.6)	0.265
Previous PCI	10 (21.7)	15 (22.7)	0.902
Final kissing balloon inflation	22 (47.8)	66 (100.0)	<0.001
Side-branch loss	3 (6.5)	0 (0.0)	0.067

Data were presented as mean ± SD or n (%). Student t-test was used for continuous variables; chi-square or Fisher's exact test was used for categorical variables, as appropriate
EF: Ejection fraction, PCI: Percutaneous coronary intervention, SD: Standard deviation

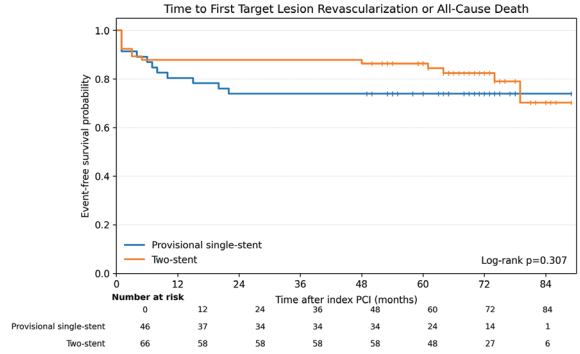

Figure 2. Kaplan-Meier event-free survival for time to first target lesion revascularization or all-cause death according to final stenting strategy. Censoring marks and numbers at risk were shown
PCI: Percutaneous coronary intervention, TLR: Target lesion revascularization

Table 2. Final procedural pathway among patients initially treated with a provisional strategy

Final procedural pathway among initially provisional cases (n=54)	n (%)
Successful single-stent result without documented SB compromise and without final kissing balloon inflation	14 (25.9)
Successful single-stent result without documented SB compromise, with final kissing balloon inflation	5 (9.3)
SB ballooning/POT-side-POT	7 (13.0)
Final kissing balloon inflation + POT	17 (31.5)
Conversion to TAP	6 (11.1)
Conversion to culotte	2 (3.7)
Side-branch loss	3 (5.6)

Data were presented as n (%). No inferential statistical comparison was performed
POT: Proximal optimization technique, SB: Side branch, TAP: T-and-protrusion

Table 3. Clinical characteristics according to the occurrence of the composite endpoint

Parameter	No endpoint (n=87)	Endpoint present (n=25)	p-value
Age, years	61.1±11.3	65.4±13.8	0.111
Main-vessel stent length, mm	22.4±7.4	23.0±7.1	0.738
Main-vessel stent diameter, mm	2.95±0.36	3.08±0.56	0.179
Bifurcation angle, degrees	53.3±15.3	54.5±16.8	0.741
EF, %	49.4±9.0	42.6±13.6	0.004
Diabetes mellitus	25 (28.7)	13 (52.0)	0.030
Hypertension	65 (74.7)	15 (60.0)	0.151
Chronic kidney disease	12 (13.8)	5 (20.0)	0.526
Previous coronary artery disease	34 (39.1)	9 (36.0)	0.780
Multivessel disease	50 (57.5)	13 (52.0)	0.627
Previous PCI	20 (23.0)	5 (20.0)	0.752
Final kissing balloon inflation	71 (81.6)	17 (68.0)	0.144

Data were presented as mean ± SD or n (%). Student t-test, chi-square test, or Fisher's exact test was used as appropriate. The composite endpoint was first TLR or all-cause death
EF: Ejection fraction, PCI: Percutaneous coronary intervention, TLR: Target lesion revascularization, SD: Standard deviation

Table 4. Univariable logistic regression analysis for the composite endpoint

Variable	Crude OR	95% CI	p-value
Two-stent strategy	0.69	0.28-1.70	0.426
Age, per 1-year increase	1.03	0.99-1.07	0.113
EF, per 5% decrease	1.34	1.08-1.67	0.007
Diabetes mellitus	2.69	1.08-6.69	0.034

Univariable binary logistic regression
CI: Confidence interval, EF: Ejection fraction, OR: Odds ratio

Table 5. Multivariable logistic regression analysis for the composite endpoint

Variable	Adjusted OR	95% CI	p-value
Two-stent strategy	0.68	0.26-1.77	0.433
EF, per 5% decrease	1.31	1.05-1.63	0.017
Diabetes mellitus	2.08	0.79-5.46	0.136

Multivariable binary logistic regression. Model n=112; composite events =25; AUC=0.688; Hosmer-Lemeshow chi-square =4.13, df=8, p=0.845; AIC=116.04; McFadden pseudo-R^2=0.092
CI: Confidence interval, EF: Ejection fraction, OR: Odds ratio, AUC: Area under the curve, AIC: Akaike information criterion

Previous studies have not identified a single strategy that is suitable for all bifurcation lesions. The extended EBC MAIN follow-up favored a stepwise provisional approach in many true left-main bifurcations, while the

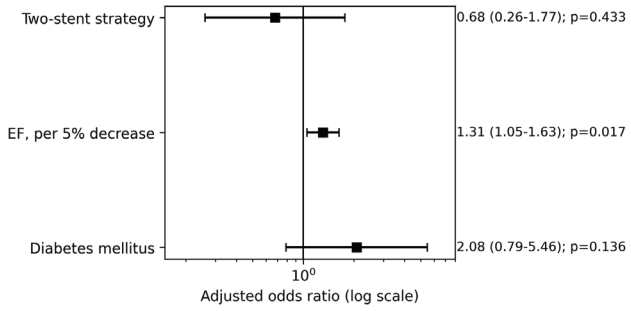

Figure 3. Forest plot of the parsimonious multivariable logistic regression model for the composite endpoint
EF: Ejection fraction

pooled DKCRUSH X analysis reported better long-term outcomes with systematic DK-crush in selected true complex lesions (6,7). Similar findings were reported in DEFINITION II and DKCRUSH-V, both of which enrolled patients with anatomically complex bifurcations (11,12). By contrast, earlier randomized trials and pooled analyses generally supported a simpler provisional strategy for many bifurcation lesions, although most of those studies included elective or mixed clinical populations (13-15). The STEMI-BIF registry added a different perspective by highlighting the potential importance of active side-branch protection during primary PCI (8). These data indicate that Medina 1.1.1 anatomy alone is not sufficient to determine treatment. Side-branch size, lesion length, ostial disease, the amount of myocardium supplied, bifurcation angle, thrombus burden, and flow after main-vessel stenting all need to be considered.

The quality of the final procedural result is also relevant. Final kissing balloon inflation was performed more often in the two-stent group, as expected; however, the present dataset did not include sufficient detail to assess the individual effect of each optimization step. Recent data from DKCRUSH VIII supported the use of intravascular imaging in complex bifurcation PCI (9). A recent systematic review also reported a lower risk of side-branch compromise with active protection techniques than with wire-only protection in selected patients (10). Intravascular imaging, quantitative side-branch measurements, and detailed information on stent expansion were not routinely available in our cohort. We, therefore, could not determine whether differences in procedural optimization influenced the association between stenting strategy and long-term outcome.

Lower EF was associated with higher odds of the composite endpoint in both univariable and multivariable analyses. This finding is clinically expected after STEMI, because reduced ventricular function reflects the extent of myocardial injury and is closely related to subsequent heart failure and mortality (16,17). Diabetes mellitus was

associated with the composite endpoint in the univariable analysis but did not remain statistically significant after adjustment. Diabetes is an established predictor of adverse outcomes after primary PCI for STEMI (18). In the present cohort, attenuation of this association after adjustment may reflect the limited number of events and reduced statistical power. Accordingly, this finding should not be interpreted as evidence that diabetes lacks prognostic importance in this population.

From a practical standpoint, our findings do not support the routine use of either a single-stent strategy or a two-stent strategy for all culprit Medina 1.1.1 lesions. A provisional approach may be appropriate when the side branch remains patent and the angiographic result is acceptable after main-vessel stenting. A second stent may be required when a relevant side branch has persistent flow limitation, severe residual ostial stenosis, or dissection. The present study was not designed to compare individual two-stent techniques or to define an optimal procedural algorithm. Instead, its findings support choosing the final strategy based on lesion anatomy, the clinical setting, and the result obtained during the procedure.

Study Limitations

This study has several limitations. Its retrospective, single-center design limits the generalizability of the findings. The stenting strategy was selected by the treating operator rather than randomly assigned; therefore, residual confounding and confounding by indication cannot be excluded. The exclusion of surviving patients with documented premature discontinuation of dual antiplatelet therapy may also have introduced selection bias. The sample size was modest, and only 25 patients reached the primary endpoint. This reduced statistical power and limited the number of variables that could be included in the multivariable model.

The two-stent group also included different techniques, which may have introduced procedural heterogeneity. Detailed measures of lesion complexity, thrombus burden, side-branch dimensions, intravascular imaging findings, coronary physiology, and stent expansion were not consistently available. Clinical outcomes were identified retrospectively, and some events may therefore have been missed or incompletely documented. Cause-specific mortality data were not consistently available; therefore, cardiovascular and non-cardiovascular deaths could not be analyzed separately. This limitation restricts the interpretation of the all-cause mortality component of the composite endpoint. Competing-risk methods were not used to analyze the first TLR, although death precluded its subsequent occurrence. Consequently, the Kaplan-Meier analysis may have overestimated the cumulative incidence of TLR, and this secondary endpoint should be interpreted cautiously. The exploratory logistic regression did not account for differences in event timing or censoring and should therefore be interpreted cautiously. The logistic regression model showed only modest discrimination, and its calibration was assessed in the same small cohort in which it was developed. External validation was not performed. The study nevertheless included a clearly defined population with culprit Medina 1.1.1 bifurcation lesions and a long-term follow-up based on time to first event.

Conclusion

In patients undergoing primary PCI for STEMI with a culprit Medina 1.1.1 bifurcation lesion, long-term event-free survival was not significantly different between those with a final single-stent result and those treated with two stents. Lower baseline EF was the variable most consistently associated with the composite endpoint. Given the retrospective design, limited sample size, and non-randomized treatment selection, these findings should be interpreted with caution and do not establish the superiority or equivalence of either strategy. In daily practice, interventional cardiologists may consider a provisional approach when the side branch remains patent and the angiographic result after main-vessel stenting is acceptable, reserving a second stent for persistent side-branch flow limitation, severe residual ostial stenosis, or dissection. Larger prospective studies are needed to clarify the long-term effects of the stenting strategy in this clinical setting.

Ethics

Ethics Committee Approval: The study was approved by the Sakarya University Local Institutional Ethics Committee (approval no: 71522473/050.01.04/617, date: 20.11.2020).

Informed Consent: The requirement for written informed consent was waived because of the retrospective, anonymized design.

Footnotes

Authorship Contributions

Surgical and Medical Practices: F.T., D.Y., Concept: F.T., Design: F.T., Data Collection or Processing: F.T., D.Y., Analysis or Interpretation: F.T., D.Y., Literature Search: F.T., D.Y., Writing: F.T.

Conflict of Interest: The authors declared that there were no conflicts of interest.

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

Artificial Intelligence (AI)-Assisted Technologies Statement: The authors declare that artificial intelligence–assisted technologies (such as large language models) were used only for language editing and grammar

improvement. The authors take full responsibility for the scientific content, data interpretation, and conclusions of the manuscript.

References

1. Byrne RA, Rossello X, Coughlan JJ, et al.; ESC Scientific Document Group. 2023 ESC guidelines for the management of acute coronary syndromes. Eur Heart J. 2023;44:3720-826.
2. Medina A, Suárez de Lezo J, Pan M. Una clasificación simple de las lesiones coronarias en bifurcación [A new classification of coronary bifurcation lesions]. Rev Esp Cardiol. 2006;59:183. Spanish.
3. Burzotta F, Lassen JF, Lefèvre T, et al. Percutaneous coronary intervention for bifurcation coronary lesions: the 15th consensus document from the European Bifurcation Club. EuroIntervention. 2021;16:1307-17.
4. Albiero R, Burzotta F, Lassen JF, et al. Treatment of coronary bifurcation lesions, part I: implanting the first stent in the provisional pathway. The 16th expert consensus document of the European Bifurcation Club. EuroIntervention. 2022;18:e362-76.
5. Lassen JF, Albiero R, Johnson TW, et al. Treatment of coronary bifurcation lesions, part II: implanting two stents. The 16th expert consensus document of the European Bifurcation Club. EuroIntervention. 2022;18:457-70.
6. Arunothayaraj S, Egred M, Banning AP, et al. Stepwise provisional versus systematic dual-stent strategies for treatment of true left main coronary bifurcation lesions. Circulation. 2025;151:612-22.
7. Chen SL, Kan J, Santoso T, et al. Double-kissing crush versus provisional stenting in patients with true coronary artery bifurcation lesions: a pooled individual patient-level analysis of randomised trials (DKCRUSH X trial). AsiaIntervention. 2025;11:178-88.
8. Güner A, Serin E, Zehir R, et al. Active side-branch protection strategy for culprit bifurcation lesion in STEMI patients: the multicenter STEMI-BIF registry. Catheter Cardiovasc Interv. 2025;106:2647-60.
9. Gao X, Kan J, Chen Y, et al.; DKCRUSH VIII Trial Group. IVUS or angiography guidance for percutaneous coronary intervention in complex coronary bifurcation lesions: the DKCRUSH VIII randomized clinical trial. J Am Coll Cardiol. 2026;87:2080-95.
10. Lee JV, Emmanuela M, Patricio A, et al. Side branch protection in coronary bifurcation stenting: a systematic review and meta-analysis of jailed wire versus jailed balloon technique. Am J Cardiol. 2026;267:58-67.
11. Zhang JJ, Ye F, Xu K, et al. Multicentre, randomized comparison of two-stent and provisional stenting techniques in patients with complex coronary bifurcation lesions: the DEFINITION II trial. Eur Heart J. 2020;41:2523-36.
12. Chen SL, Zhang JJ, Han Y, et al. Double kissing crush versus provisional stenting for left main distal bifurcation lesions: DKCRUSH-V randomized trial. J Am Coll Cardiol. 2017;70:2605-17.
13. Steigen TK, Maeng M, Wiseth R, et al; Nordic PCI Study Group. Randomized study on simple versus complex stenting of coronary artery bifurcation lesions: the Nordic bifurcation study. Circulation. 2006;114:1955-61.
14. Hildick-Smith D, de Belder AJ, Cooter N, et al. Randomized trial of simple versus complex drug-eluting stenting for bifurcation lesions: the British Bifurcation Coronary Study: old, new, and evolving strategies. Circulation. 2010;121:1235-43.
15. Behan MW, Holm NR, Curzen NP, et al. Simple or complex stenting for bifurcation coronary lesions: a patient-level pooled-analysis of the Nordic Bifurcation Study and the British Bifurcation Coronary Study. Circ Cardiovasc Interv. 2011;4:57-64.
16. Halkin A, Singh M, Nikolsky E, et al. Prediction of mortality after primary percutaneous coronary intervention for acute myocardial infarction: the CADILLAC risk score. J Am Coll Cardiol. 2005;45:1397-405.
17. Lei Z, Li B, Li B, Peng W. Predictors and prognostic impact of left ventricular ejection fraction trajectories in patients with ST-segment elevation myocardial infarction. Aging Clin Exp Res. 2022;34:1429-38.
18. Dziewierz A, Zdzierak B, Malinowski KP, et al. Diabetes mellitus is still a strong predictor of periprocedural outcomes of primary percutaneous coronary interventions in patients presenting with ST-segment elevation myocardial infarction (from the ORPKI Polish National Registry). J Clin Med. 2022;11:6284.

Original Article

DOI: 10.4274/haseki.galenos.2026.63634
Med Bull Haseki 2026;64(4):244-253

Diagnostic Accuracy of the Dietary Screening Tool for Detecting Malnutrition in Hospitalized Older Adults: Comparison with the Mini Nutritional Assessment-1999 and Nutritional Risk Screening-2002

Halime Selen[1,2], Zeynep Ertekin[1,3]

[1]*Agri Ibrahim Cecen University Faculty of Health Sciences, Department of Nutrition and Dietetics, Agri, Türkiye*
[2]*Mardin Artuklu University Faculty of Health Sciences, Department of Nutrition and Dietetics, Mardin, Türkiye*
[3]*Agri Training and Research Hospital, Clinic of General Surgery, Agri, Türkiye*

Abstract

Aim: Malnutrition is a common and clinically important problem among hospitalized older patients, and accurate screening tools are essential for the early identification of nutritional risk. The aim of this study was to determine the prevalence of malnutrition in hospitalized older patients and to evaluate the compatibility of the dietary screening tool (DST) for older adults with the Mini Nutritional Assessment-1999 (MNA-1999) and the Nutritional Risk Screening-2002 (NRS-2002).

Methods: This single-center analytical cross-sectional study was conducted between January 1 and March 31, 2025, and included 302 older adults aged 65-87 years. Participants were assessed using a questionnaire that included general information, the DST, MNA-1999, and NRS-2002. Data were analyzed using descriptive statistics, the chi-square test, Pearson correlation analysis, and parametric and non-parametric tests. The diagnostic performance of the DST was evaluated using receiver operating characteristic analysis [area under the curve (AUC), sensitivity, specificity, and cut-off values].

Results: According to the DST, MNA-1999, and NRS-2002, the prevalence of malnutrition was 5%, 40.7%, and 32.8%, respectively. Advanced age, low body mass index, reduced mid-upper arm and calf circumferences, greater weight loss during the previous three months, prolonged hospital stay, and hospitalization in hematology or palliative care units were significantly associated with malnutrition according to the DST (all $p<0.001$), whereas sex was not. Using MNA-1999 as the reference standard, the DST demonstrated a sensitivity of 50.4%, specificity of 86.6%, positive predictive value of 72%, and negative predictive value of 71.8%. In contrast, NRS-2002 showed a sensitivity of 27.6%, specificity of 5.6%, positive predictive value of 16.7%, and negative predictive value of 10.1%. Receiver operating characteristic analysis showed good discriminative ability against MNA-1999 [AUC=0.851, 95% confidence interval (CI): 0.808-0.893] and acceptable performance against NRS-2002 (AUC=0.779, 95% CI: 0.729-0.830). The optimal DST cut-off value was 51.5 for both reference standards. Moderate correlations were observed between DST and MNA-1999 ($r=0.694$) and between DST and NRS-2002 ($r=-0.637$), whereas NRS-2002 showed a strong negative correlation with MNA-1999 ($r=-0.842$) (all $p<0.001$).

Conclusion: Healthcare professionals should exercise caution when using the DST as a malnutrition screening tool. Moreover, the applicability of the DST in different populations needs to be further evaluated.

Keywords: Dietary screening tool, hospitalization, malnutrition, Mini Nutritional Assessment-1999, nutrition assessment, Nutritional Risk Screening-2002

Corresponding Author: Assist. Prof. Halime Selen, Agri Ibrahim Cecen University Faculty of Health Sciences, Department of Nutrition and Dietetics, Agri, Türkiye; Mardin Artuklu University Faculty of Health Sciences, Department of Nutrition and Dietetics, Mardin, Türkiye
E-mail: halimeselen@gmail.com **ORCID:** orcid.org/0000-0002-3705-0875
Received: 18.07.2025 **Accepted:** 17.07.2026 **Publication Date:** 24.09.2026

Cite this article as: Selen H, Ertekin Z. Diagnostic accuracy of the dietary screening tool for detecting malnutrition in hospitalized older adults: comparison with the Mini Nutritional Assessment-1999 and Nutritional Risk Screening-2002. Med Bull Haseki. 2026;64(4):244-253

Introduction

The population of Europe is progressively aging. Recent statistics as of January 1, 2025, indicate that over one-fifth (22.0%) of the European population is aged 65 years and older (1). Advancements in health technologies, including diagnostics, medical devices, and pharmaceutical treatments, have contributed to the extension of life expectancy. Further developments in medical care, such as improved control of infectious diseases, enhanced hemodynamic management, and advances in mechanical ventilation and intensive care units, have also contributed to prolonging life expectancy (2). However, studies have indicated that a longer hospital stay is associated with a heightened risk of malnutrition (3,4). A recent study reported an incidence of hospital-acquired malnutrition of 1% among patients hospitalized for more than 14 days (5). The same study also identified hospital-acquired malnutrition as being associated with prolonged hospital stays, cognitive impairment, and pressure sores (5). A meta-analysis evaluating the prevalence of malnutrition among older adults aged 65 years and over in European countries reported a prevalence of 28.0% among hospitalized older patients (6). A recent study conducted in 10 geriatric centers in Spain found that 13.4% of participants were classified as malnourished, while 47.4% were at risk of malnutrition (7). Globally, the prevalence of malnutrition among older adults has been reported as 18.6% (8).

Malnutrition remains a frequent issue among hospitalized older patients. It increases patient morbidity and mortality, leading to substantial healthcare costs (9). Therefore, screening for malnutrition and providing nutritional support are essential components of patient management (9). Various screening tools recommended by scientific organizations are available to assess patients' risk of malnutrition (10-13). The European Society for Parenteral and Enteral Nutrition (ESPEN), followed by the Turkish Society of Clinical Enteral and Parenteral Nutrition, recommends the use of the Mini Nutritional Assessment-1999 (MNA-1999) screening tool for older patients and the Nutritional Risk Screening-2002 (NRS-2002) for hospitalized patients (10). Recently, a new screening tool called the dietary screening tool (DST) has also been proposed to assess the nutritional status of older adults (14). However, only a limited number of studies have evaluated the effectiveness of the DST.

Although the DST has been developed for nutritional screening in older adults, evidence regarding its diagnostic performance in hospitalized older patients remains limited. Moreover, direct comparisons of the DST with the two most widely recommended nutritional screening tools, the MNA-1999 and the NRS-2002, are scarce. Therefore, the aim of this study was to determine the prevalence of malnutrition among hospitalized older adults and to evaluate the diagnostic performance, agreement, and optimal cut-off value of the DST using MNA-1999 and NRS-2002 as reference standards. We hypothesized that the DST would demonstrate significant agreement with MNA-1999 and NRS-2002 and acceptable diagnostic performance for identifying malnutrition risk in hospitalized older adults.

Materials and Methods

Compliance with Ethical Standards

To ensure the ethical significance of the study, approval was obtained from the Scientific Research Ethics Committee of Agri Ibrahim Cecen University with (approval no.: 384, date: 31.10.2024). In addition, the study was conducted in accordance with the Declaration of Helsinki.

Study Period, Design, Location, Population, and Sample

This was a single-center analytical cross-sectional study conducted among hospitalized older patients between January 1, 2025, and March 31, 2025. The research took place in the inpatient service units of Agri Training and Research Hospital, located in Agri province, Türkiye. Verbal and written informed consent were obtained from all participants. The study population and sample comprised older adults aged 65 and above who were hospitalized in the inpatient units of Agri Training and Research Hospital from January 1 to March 31, 2025. The sample size was based on data collected from all patients who met the inclusion criteria during the three-month period.

Inclusion Criteria

The inclusion criteria for the study included patients aged 65 years and above who possessed intact communication abilities, demonstrated normal cognitive function, and provided informed consent to participate.

Data Collection Tools

Data were collected using the "General Information Questionnaire," "DST," "MNA-1999," and "NRS-2002" tools.

General Information Questionnaire: The data collection instrument, developed by the researchers based on an extensive review of relevant literature, encompassed demographic variables such as age and sex, anthropometric measurements, and information regarding the duration of hospital stay and the clinical department of admission. Body weight (kg) and height (m) were measured. [Body mass index (BMI)=kg/m^2] was calculated based on the participants' body weight and height measurements and subsequently categorized in accordance with World Health Organization standards: underweight (<18.5 kg/m^2), normal weight (18.5-24.9 kg/m^2), overweight (25.0-29.9 kg/m^2), and obese (≥30 kg/m^2) (15).

Mid-upper arm circumference was measured with the participant's elbow flexed at 90°, using a measuring tape placed at the midpoint between the acromial process of the scapula and the olecranon process of the ulna (16). Calf circumference was measured according to the method described by Erdoğan et al. (17).

Dietary Screening Tool: The DST was originally developed by Bailey et al. (14) as a dietary screening instrument for older adults. Its validity and reliability for the Turkish population were established by Toklu Baloğlu (18). The DST consists of 25 questions evaluated out of a total of 105 points (18). Based on the DST score, participants were classified into three groups: those in the lowest 25th percentile were categorized as "at risk," those between the 25th and 75th percentiles as "potentially at risk," and those in the highest 25th percentile as "not at risk" (18).

The Food and Nutrition Photo Catalogue: Measurements and quantities was used to help participants estimate the quantities and portion sizes of the food groups included in the DST (19). Portion sizes were evaluated according to the food portion guidelines provided in the Turkish Nutrition Guidelines 2022 and were used in calculating the DST score (20).

Mini Nutritional Assessment-1999: It was developed by Guigoz et al. (21) to assess malnutrition in older adults receiving home care, residing in nursing homes, or being hospitalized. Its validity and reliability for the Turkish population were established by Sarıkaya (22). The tool includes assessments of anthropometric measurements, general evaluation, a brief dietary assessment, and subjective self-perceptions. The MNA-1999 consists of two main parts: a screening (21,22) and an assessment section. If the score obtained in the initial screening section is ≤11, the assessment section is administered in order to calculate the total score. A total score of <17 indicates the "presence of malnutrition" (14,18); a score between 17 and 23.5 indicates a "risk of malnutrition," and a score ≥23.5 reflects "adequate nutritional status" (21,22). In the present study, since all patients had an MNA-1999 screening (21,22) ≤11, the assessment section was completed for all participants, and final classifications were made based on the total scores.

Nutritional Risk Screening-2002: It was developed by Kondrup et al. (23) to assess the nutritional status of hospitalized patients, and its validity and reliability for the Turkish population were established by Bolayır (24). The tool comprises four components that assess nutritional status based on weight loss, dietary intake, BMI, disease severity, and age. An additional 1 point is assigned to patients aged ≥70 years. Based on the total score, patients are classified as "not at nutritional risk" if they score <3 points and "at nutritional risk" if they score ≥3 points (23,24).

Statistical Analysis

The data obtained in the study were analyzed using SPSS software version 27.0 (IBM Corp., Armonk, NY, U.S.A.). Descriptive statistical methods were used to summarize the data. The normality of data distribution was assessed through Q-Q plots, and normal distribution was considered present when skewness and kurtosis values fell within the ±3 range. For normally distributed data, the relationships between continuous variables were examined using Pearson correlation analysis. Associations between categorical variables were analyzed using the chi-square test. Differences among more than two independent groups were assessed using one-way analysis of variance for normally distributed variables and the Kruskal-Wallis H test for non-normally distributed variables. The discriminatory power of DST was evaluated using receiver operating characteristic (ROC) analysis; diagnostic performance metrics, including area under the curve (AUC), sensitivity, specificity, and cut-off values with corresponding diagnostic measures, were calculated. A $p<0.05$ value was considered statistically significant.

Results

The baseline characteristics of the older adult participants are presented in Table 1. A total of 302 individuals aged 65-87 years (mean age: 70.22±3.48 years) were included in the study, of whom 54.3% were male and 45.7% were female. Most patients were within the normal BMI range. Most hospitalizations lasted 4-5 days, and the most common admitting departments were general surgery, pulmonology, and orthopedics.

The prevalence of malnutrition according to the characteristic (ROC) analysis is shown in Table 2. According to DST, 28.5% of patients were classified as high-risk and 47.0% as potentially at risk. Based on MNA-1999, 40.7% of participants were malnourished and 42.4% were at risk of malnutrition. According to NRS-2002, 32.8% of patients were at nutritional risk.

Table 3 presents a comparison of categorical variables according to DST classification. Significant associations were observed between DST categories and age group, current BMI, length of hospital stay, hospitalization department, and BMI three months prior (all $p<0.001$). In contrast, sex was not significantly associated with DST classification ($p=0.422$).

Comparisons of continuous variables across DST categories are presented in Table 4. Patients in the risk group were significantly older and had lower BMI and smaller mid-upper arm and calf circumferences compared with those in other groups ($p>0.05$). In addition, individuals in the risk group had longer hospital stays and greater percentage weight loss in the preceding three months ($p<0.001$).

Table 1. Baseline characteristics of the study population (n=302)

Variables		n	%
Age (years) (mean ± SD = 70.22±3.48)	<75	284	94.0
	≥75	18	6.0
Sex	Female	138	45.7
	Male	164	54.3
Current BMI	Underweight	7	2.3
	Normal weight	250	82.8
	Overweight	43	14.2
	Obese	2	0.7
BMI values from three months prior	Normal weight	230	76.2
	Overweight	67	22.2
	Obese	5	1.7
Length of hospital stay (days)	≤3	38	12.6
	4-5	200	66.2
	≥6	64	21.2
Hospitalization department	General surgery	85	28.1
	Internal medicine	30	9.9
	Pulmonology	70	23.2
	Orthopedics	47	15.6
	Palliative	21	7.0
	Hematology	30	9.9
	Oncology	19	6.3

Descriptive statistics were expressed as mean ± standard deviation or frequency (n) and percentage (%)
BMI: Body mass index, SD: Standard deviation

The diagnostic performance of DST and NRS-2002, using MNA-1999 as the reference standard, is presented in Table 5. Dietary screening tool showed a sensitivity of 50.4%, a specificity of a greater positive predictive value of 72.1%, a negative predictive value of 72.1% (p<0.001) and a value of 71.8%. In contrast, specificity demonstrated a positive value of 27.6%, a specificity of 5, a positive predictive value of 16.7%, and a negative predictive value of 10.1%.

Table 2. Malnutrition prevalence of patients according to three different malnutrition screening tools (n=302)

Malnutrition screening tools	Score	n	%
DST	At risk	86	28.5
	Potentially at risk	142	47.0
	Not at risk	74	24.5
MNA-1999	Malnutrition	123	40.7
	At risk of malnutrition	128	42.4
	Normal nutritional status	51	16.9
NRS-2002	No risk of malnutrition	203	67.2
	Risk of malnutrition	99	32.8

Descriptive statistics are expressed as frequency (n) and percentage (%)
DST: Dietary screening tool, MNA-1999: Mini Nutritional Assessment-1999, NRS-2002: Nutritional Risk Screening-2002

Table 3. Comparison of selected categorical variables across the study population classified according to DST (n=302)

		DST						Chi-square	p
		At risk		Potentially at risk		Not at risk			
		n	%	n	%	n	%		
Age groups (years)	<75	73	84.9	138	97.2	73	98.6	15.221[FE]	<0.001
	≤75	12	15.1	4	2.8	1	1.4		
Sex	Female	40	46.5	69	48.6	29	39.2	1.761[P]	0.422
	Male	46	53.5	73	51.4	45	60.8		
Current BMI	Underweight	5	5.8	2	1.4	0	0.0	41.948[FE]	<0.001
	Normal weight	78	90.7	126	88.7	46	62.2		
	Overweight	3	3.5	13	9.2	27	36.4		
	Obese	0	0.0	1	0.7	1	1.4		
Length of hospital stay (days)	≤3	9	10.5	13	9.2	16	21.6	26.928[P]	<0.001
	4-5	45	52.3	103	72.5	52	70.3		
	≥6	32	37.2	26	18.3	6	8.1		
Hospitalization department	General surgery	20	23.25	38	26.8	27	36.5	57.211[FE]	<0.001
	Internal medicine	7	8.1	12	8.4	11	14.9		
	Pulmonology	14	16.3	31	21.8	25	33.8		
	Orthopedics	14	16.3	24	16.9	9	12.1		
	Palliative	8	9.3	13	9.2	0	0.0		
	Hematology	20	23.25	10	7.0	0	0.0		
	Oncology	3	3.5	14	9.9	2	2.7		
BMI categories from three months prior	Normal weight	77	89.5	115	81.0	38	51.3	33.612[FE]	<0.001
	Overweight	9	10.5	25	17.6	33	44.6		
	Obese	0	0.0	2	1.4	3	4.1		

Descriptive statistics are expressed as frequency (n) and percentage (%). [FE]=Fisher Exact chi-square. [P]: Pearson chi-square
BMI: Body mass index, DST: Dietary screening tool

Table 4. Comparison of general characteristics across the study population classified according to DST (n=302)									
	DST	**Mean**	**SD**	**Q25**	**Q50**	**Q75**	**Test value**	**p**	**Post-hoc**
Age (years)	At risk (1)	71.88	4.14	70.00	72.00	73.00	34.497[KW]	<0.001	1>3 2>3
	Potentially at risk (2)	70.06	2.81	69.00	70.00	72.00			
	Not at risk (3)	68.58	2.94	66.00	70.00	70.00			
Current BMI	At risk (1)	20.77	1.72	19.50	20.70	21.30	44.551[F]	<0.001	3>1 3>2
	Potentially at risk (2)	21.98	2.34	20.70	21.60	22.90			
	Not at risk (3)	24.17	2.68	22.30	23.50	26.20			
Hospital length of stay (days)	At risk (1)	5.45	1.91	4.00	5.00	7.00	15.857[F]	<0.001	1>2 1>3 2>3
	Potentially at risk (2)	4.64	1.25	4.00	5.00	5.00			
	Not at risk (3)	4.09	1.14	4.00	4.00	5.00			
MUAC (cm)	At risk (1)	22.08	2.30	21.00	22.00	23.00	89.854[F]	<0.001	3>1 3>2 2>1
	Potentially at risk (2)	24.85	4.18	22.00	24.00	27.00			
	Not at risk (3)	30.07	4.85	27.00	28.50	34.00			
Calf circumference (cm)	At risk (1)	29.63	2.31	29.00	30.00	31.00	90.830[F]	<0.001	3>1 3>2 2>1
	Potentially at risk (2)	32.71	4.19	30.00	31.00	34.00			
	Not at risk (3)	37.84	5.10	34.00	36.00	42.00			
Percentage weight loss over the preceding three months (%)	At risk (1)	7.65	3.37	5.00	7.90	10.00	70.178[F]	<0.001	1>2 1>3 2>3
	Potentially at risk (2)	6.06	2.97	4.00	6.00	8.00			
	Not at risk (3)	2.99	2.07	0.00	3.15	4.60			
BMI values from three months prior	At risk (1)	22.47	1.64	21.40	22.15	23.20	24.757[F]	<0.001	3>1 3>2 2>1
	Potentially at risk (2)	23.47	2.08	22.10	23.10	24.20			
	Not at risk (3)	24.92	2.69	22.80	24.55	26.70			

Descriptive statistics are expressed as mean ± standard deviation. [KW]: Kruskal-Wallis H test. [F]=ANOVA
BMI: Body mass index, DST: Dietary screening tool, MUAC: Mid-upper arm circumference, ANOVA: Analysis of variance

Table 5. Comparison of the diagnostic accuracy (sensitivity, specificity, PPV, NPV) of three screening tools using MNA-1999 as the reference standard								
		MNA-1999				**Chi-square**	**p**	**Phi coefficient**
		Malnutrition		**Normal nutritional status**				
		n	**%**	**n**	**%**			
DST	At risk	62	50.4	24	13.4	48.999	<0.001	0.403
	Not at risk	61	49.6	155	86.6			
NRS-2002	No risk of malnutrition	34	27.6	169	94.4	147.507	<0.001	-0.699
	Risk of malnutrition	89	72.4	10	5.6			

DST: Dietary screening tool, MNA-1999: Mini Nutritional Assessment-1999, NRS-2002: Nutritional Risk Screening-2002

Figure 1 presents the ROC curves demonstrating the diagnostic performance of the DST, using MNA-1999 (a) and NRS-2002 (b) as reference standards. Specificity analysis based on positive 9, the AUC was 0.851 [standard deviation (SD)=0.022; 95% negative 0.808-0.893; p<0.001], indicating that the DST demonstrated good, statistically significant discriminatory performance for normal nutritional status. According to the NRS-2002-based analysis, the AUC was 0.779 [SD=0.026; 95% confidence interval (CI): 0.729-0.830; p<0.001], suggesting that the DST demonstrated acceptable and statistically significant discriminatory performance in identifying individuals without risk of malnutrition. In both analyses, the CIs above 0.5 indicate consistent discriminative ability. The optimal cut-off value (NRS-2002-based) was determined using the Youden index. When MNA-1999 was used as the reference standard, the highest Youden index was 0.622 at a cut-off value of 51.5, corresponding to a sensitivity of 68.7% and a specificity of 93.5%, indicating the best balance between sensitivity and specificity. Similarly, when NRS-2002 was used as the reference standard, the highest Youden index (0.540) was

obtained at the cut-off value of 51.5, with a sensitivity of 61.1% and a specificity of 92.9%. Although higher cut-off values (e.g., 57.5) produced greater positive likelihood ratios due to increased specificity, they were associated with substantially lower sensitivity. Therefore, a DST cut-off value of 51.5 appears to provide the most appropriate overall diagnostic performance for identifying malnutrition risk in hospitalized older patients.

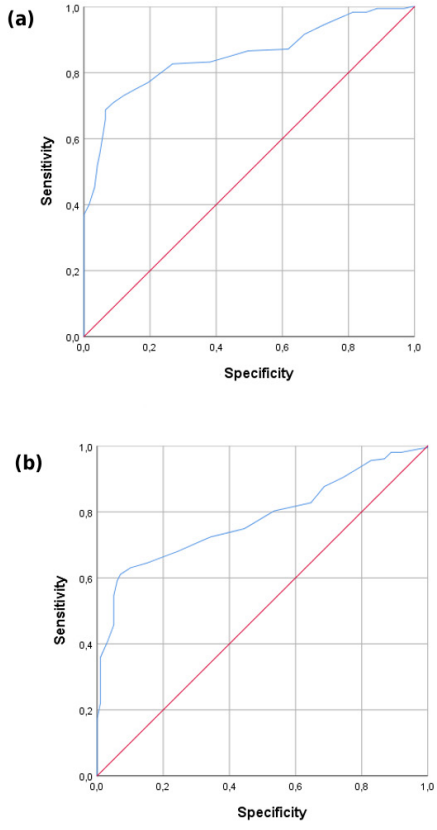

Figure 1. ROC curves demonstrating the diagnostic performance of DST against MNA-1999 (a) and NRS-2002 (b) as reference standards
ROC: Receiver operating characteristic, DST: Dietary screening tool, MNA-1999: Mini Nutritional Assessment-1999, NRS-2002: Nutritional Risk Screening-2002

Table 6. Concordance analysis among the three screening tools

		DST	MNA-1999	NRS-2002
DST	r	1		
	p			
MNA-1999	r	0.694	1	
	p	<0.001		
NRS-2002	r	-0.637	-0.842	1
	p	<0.001	<0.001	

r: Pearson correlation coefficient
DST: Dietary screening tool, MNA-1999: Mini Nutritional Assessment-1999, NRS-2002: Nutritional Risk Screening-2002

The correlation analysis between the screening tools is shown in Table 6. A moderate positive correlation was found between DST and MNA-1999 (r=0.694, p<0.001), whereas a moderate negative correlation was found between DST and NRS-2002 (r=-0.637, p<0.001). A strong negative correlation was observed between NRS-2002 and MNA-1999 (r=-0.842, p<0.001).

Discussion

Despite advances in health technologies that have contributed to increased life expectancy, malnutrition remains a significant issue among hospitalized older adults. This study evaluated the prevalence of malnutrition in hospitalized older adults and assessed the diagnostic performance and agreement of the DST compared with the MNA-1999 and NRS-2002.

In the present study, between 28.5% and 40.7% of patients were at risk of malnutrition based on the three screening tools used (Table 2). Similarly, a meta-analysis conducted across European countries, which evaluated the outcomes of 22 different screening tools in individuals aged 65 and over, reported rates of malnutrition risk ranging from 17.5% to 40.6% in hospital, nursing home, and community-based screenings (6). In a prospective cohort study conducted in Ireland involving patients aged 70 years and older and utilizing the short form of the MNA-1999, 18% of the patients were identified as malnourished, and 45% were found to be at risk of malnutrition (25). In another study conducted across 25 European countries, the average prevalence of malnutrition among hospitalized patients was reported to be 12.9% (26). In a study conducted in Tehran using the MNA-1999 to assess the nutritional status of older patients, only 41.9% of individuals were found to have a normal nutritional status (27). In the United States, the prevalence of malnutrition among hospitalized patients was reported to be 8.6%, varying in severity (28). A recent study conducted in Istanbul, Türkiye, in which older patients were assessed using the MNA-1999, found that 50% of the participants were at risk of malnutrition, while 8% were already malnourished (29). The varying prevalence of malnutrition observed in the current study suggests that the choice of screening instrument may substantially influence the identification of at-risk patients in clinical settings. The higher prevalence observed, particularly with the MNA-1999, may be explained by its more comprehensive structure, which includes anthropometric, dietary, and functional parameters, potentially allowing earlier detection of nutritional decline. The prevalence observed in our cohort appears higher than that reported in studies conducted in Europe and the

United States. This difference may be related to several factors, including differences in healthcare systems, nutritional support practices, patient case mix (e.g., higher proportion of severe or chronic conditions), and possible underutilization of routine nutritional screening in certain settings. These findings highlight the need for more systematic and standardized nutritional assessment approaches in hospitalized older adult populations.

According to the DST classification used in our study, the following factors were identified as associated with a higher risk of developing malnutrition: advanced age, low BMI, reduced mid-upper arm circumference, reduced calf circumference, significant weight loss over the previous three months, prolonged hospital stay, and hospitalization in hematology or palliative care units. However, sex was not a significant factor (Tables 3 and 4). In the United States, the DST was utilized as a tool for NRS and the effectiveness of nutritional education programs was investigated (30). In a study involving individuals aged 60 and above who participated in community-based nutrition education and physical activity programs, DST assessments indicated that 2 participants were classified as being at high nutritional risk, while 53.7% were in the possible risk group (31). In a study conducted by Bailey et al. (32) investigating dietary intake among older adults, individuals classified as being at risk according to the DST were found to have significantly lower protein intake and higher total and saturated fat intake compared to those in the possible risk and not-at-risk groups (32). In a study involving patients with a prevalence, individuals in the malnourished group (n=70) demonstrated significantly lower values for body weight, mid-upper arm circumference, waist circumference, waist-to-height ratio, serum albumin, total serum cholesterol, and blood hemoglobin compared to the non-malnourished group (n=82) ($p<0.05$); however, no statistically significant difference was observed between the groups' case mix (e.g., age) ($p>0.05$) (33). In a study conducted (under specific conditions), researchers found that older adults with sarcopenia had significantly longer hospital stays compared to those without sarcopenia ($p<0.05$); however, no significant differences were observed based on sex ($p<0.05$) (34). The prolonged hospital stay, observed particularly among individuals at high-risk of malnutrition, underscores the critical importance of nutritional support in clinical management. A meta-analysis that included only longitudinal studies investigating malnutrition risk factors in older adults identified advanced age as a significant determinant of malnutrition risk, while sex was not found to be a distinguishing factor in this context (35). Similarly, several studies have reported a higher prevalence of malnutrition among older adults hospitalized in hematology and palliative care units (36,37). The current literature indicates that malnutrition is highly prevalent among older adults and is closely associated with factors such as low BMI, weight loss, and disease burden (38). Additionally, it has been reported that as the risk of malnutrition increases, so do hospital length of stay, complication rates, and mortality (39). A recent study conducted in Türkiye indicates that the prevalence of malnutrition among older adults can reach up to 50% depending on the method used and is associated with food insecurity (40). Given that the early use of screening tools in hospitalized older adults can improve clinical outcomes (41), integrating these tools into routine assessment processes could facilitate the early identification of at-risk individuals and the initiation of appropriate interventions at the earliest possible stage, thereby contributing to the reduction of morbidity and mortality. Collectively, these findings underscore the clinical importance of early ($p<0.05$) nutrition risk screening and timely, targeted nutritional intervention ($p<0.05$) in older adults, both of which can contribute to shorter hospital stays, reduced complication rates, and decreased mortality.

When comparing the diagnostic accuracy of DST and NRS-2002 with reference to the MNA-1999, MNA-1999 demonstrated moderate concordance with DST and very strong concordance with NRS-2002 (Tables 5 and 6, Figure 1). In a study population, 59.7% of individuals were classified as "at risk" according to the DST, whereas only 40.7% were found to be malnourished based on the MNA-1999. Moreover, no significant association was identified between the total scores of the DST and MNA-1999 (18). In a separate study evaluating patients with Paronlyn's disease or acquired brain injury, a positive correlation was observed between DST scores and general activity scores ($r=0.697$, $p=0.012$) (42). A study conducted in Istanbul among cancer patients demonstrated moderate concordance between MNA-1999 and NRS-2002 (43). These two instruments are the primary screening tools recommended by ESPEN (10). However, compared with MNA-1999 and NRS-2002, the DST is less commonly used for malnutrition screening. In the present study, the DST exhibited moderate concordance with both the MNA-1999 and NRS-2002. The relatively low sensitivity but high specificity of the DST obs (36,37). This study suggests that, while the tool may be effective in confirming malnutrition risk, it may fail to identify a substantial proportion of at-risk individuals. One possible explanation is that the DST primarily focuses on dietary patterns rather than clinical and functional parameters, which are more comprehensively captured by tools such as the MNA-1999. Therefore, our findings suggest that DST should be considered a complementary, rather than a standalone, screening tool in hospitalized older adult populations. While its relatively high specificity

may support confirmation of nutritional risk, its limited sensitivity indicates that relying solely on DST could result in under-identification of vulnerable patients. This study provides evidence on the clinical performance of the DST in hospitalized older adults, a population for which validation data are currently limited. Our findings, derived from direct comparison of DST with both MNA-1999 and NRS-2002 within the same patient cohort and from identification of an optimal DST cut-off value, provide practical information for clinicians regarding the strengths and limitations of DST in routine hospital practice. These findings may help guide the selection of appropriate nutritional screening strategies for hospitalized older patients.

Study Limitations

The present study has several limitations. First, it was conducted at a single center with a relatively small sample size, which may restrict the generalizability of the findings. In addition, only patients who were able to communicate effectively and who had preserved cognitive function were included in the study. This eligibility criterion may have introduced selection bias by excluding frailer, cognitively impaired older adults, who may be at even higher risk of malnutrition, thereby limiting the generalizability of the findings to the overall hospitalized older adult population. Second, body composition parameters such as muscle and adipose tissue mass were not evaluated using objective methods such as bioelectrical impedance analysis. Third, biochemical nutritional markers common to assess malnutrition, including albumin and prealbumin, were not available. Additionally, socioeconomic and familial factors that might indirectly influence nutritional status were not assessed.

Despite these limitations, this study has important strengths. To our knowledge, this is the first study to simultaneously compare the diagnostic performance and concordance of DST with both MNA-1999 and NRS-2002 in hospitalized older adults. Furthermore, the use of multiple validated nutritional screening tools within the same patient population provides a comprehensive comparison and contributes valuable data on the clinical applicability of DST in hospitalized older adults. These findings extend the current evidence regarding the clinical utility of DST and provide practical guidance for its use in hospital-based nutrition comparisons of older adults.

Conclusion

Malnutrition affected nearly one-third of hospitalized older adults in this study, highlighting the importance of routine nutritional screening in this population. Although DST showed good overall discriminative ability and moderate agreement with MNA-1999, its relatively low sensitivity suggests that relying solely on DST may result in underdiagnosis of malnutrition. Accordingly, DST should not replace established nutritional assessment tools but may serve as a complementary screening instrument. Additional prospective multicenter validation studies are needed before its routine use in hospitalized older adults can be recommended.

Ethics

Ethics Committee Approval: To ensure the ethical conduct of the study, approval was obtained from the Scientific Research Ethics Committee of Agri Ibrahim Cecen University with (approval no.: 384, date: 31.10.2024).

Informed Consent: Informed consent was obtained from all subjects involved in the study.

Footnotes

Authorship Contributions

Concept: H.S., Design: H.S., Data Collection or Processing: H.S., Z.E., Analysis or Interpretation: H.S., Literature Search: H.S., Z.E., Writing: H.S., Z.E.

Conflict of Interest: The authors declared that there were no conflicts of interest.

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

References

1. Eurostat. Population structure and ageing. 2026. Available from:s URL:https://ec.europa.eu/eurostat/statistics- explained/index.php?title=Population_structure_and_ageing&oldid=549185. Accessed July 18, 2025.
2. Wamble DE, Ciarametaro M, Dubois R. The effect of medical technology innovations on patient outcomes, 1990-2015: results of a physician survey. J Manag Care Spec Pharm. 2019;25:66-71.
3. Yılmaz Ferhatoğlu S, Dönmez NF. The effect of nutritional status on length of hospital stay in adult patients undergoing elective orthopedic surgery: a prospective analysis. Med Bull Haseki. 2020;58:228-33.
4. Cruz-Jentoft AJ, Volkert D. Malnutrition in older adults. N Engl J Med. 2025;392:2244-55.
5. Woodward T, Josephson C, Ross L, et al. A retrospective study of the incidence and characteristics of long-stay adult inpatients with hospital-acquired malnutrition across five Australian public hospitals. Eur J Clin Nutr. 2020;74:1668-76.
6. Leij-Halfwerk S, Verwijs MH, van Houdt S, et al.; MaNuEL Consortium. Prevalence of protein-energy malnutrition risk in European older adults in community, residential and hospital settings, according to 22 malnutrition screening tools validated for use in adults ≥65 years: a systematic review and meta-analysis. Maturitas. 2019;126:80-9.
7. Ruperto M, Ongan D, Josa E, Tsagari A. Assessment of hidden nutritional burden: high prevalence of disease-related malnutrition in older adults without cognitive impairment living in nursing homes in madrid-a multicentre study. Nutrients. 2025;17:3325.

8. Salari N, Darvishi N, Bartina Y, Keshavarzi F, Hosseinian-Far M, Mohammadi M. Global prevalence of malnutrition in older adults: a comprehensive systematic review and meta-analysis. Public Health Pract (Oxf). 2025;9:100583.
9. Bellanti F, Lo Buglio A, Quiete S, Vendemiale G. Malnutrition in hospitalized old patients: screening and diagnosis, clinical outcomes, and management. Nutrients. 2022;14:910.
10. Kondrup J, Allison SP, Elia M, et al.; Educational and Clinical Practice Committee, European Society of Parenteral and Enteral Nutrition (ESPEN). ESPEN guidelines for nutrition screening 2002. Clin Nutr. 2003;22:415-21.
11. House M, Gwaltney C. Malnutrition screening and diagnosis tools: Implications for practice. Nutr Clin Pract. 2022;37:12-22.
12. Jensen GL, Cederholm T, Correia MITD, et al. GLIM criteria for the diagnosis of malnutrition: a consensus report from the global clinical nutrition community. J Parenter Enter Nutr. 2019;43:32-40.
13. Skipper A, Ferguson M, Thompson K, Castellanos VH, Porcari J. Nutrition screening tools: an analysis of the evidence. JPEN J Parenter Enteral Nutr. 2012;36:292-8.
14. Bailey RL, Mitchell DC, Miller CK, et al. A dietary screening questionnaire identifies dietary patterns in older adults. J Nutr. 2007;137:421-6.
15. World Health Organization. Everyday actions for better health – WHO recommendations. 2025. Available from:s URL:https://www.who.int/europe/news-room/fact-sheets/item/a-healthy-lifestyle—who-recommendations. Accessed July 18, 2025.
16. Aksoydan E. Yaşlı ve beslenme. T.C. Sağlık Bakanlığı; Ankara. 2012.
17. Erdoğan T, Çatıkkaş NM, Kılıç C, Karan MA, Bahat G. Turkish calf circumference cut-offs derived from normative values of young reference population. Eur J Geriatr Gerontol. 2022;4:103-7.
18. Toklu Baloğlu H. Adaptation of the dietary screening tool in the elderly into Turkish and evaluation of related factors (thesis). Kayseri: Erciyes Üniv; 2023.
19. Rakıcıoğlu N, Acar Tek N, Ayaz A, Pekcan G. Food and nutrition photo catalogue: measurements and quantities. Ata Ofset Printing, Ankara; 2012.
20. Türkiye Beslenme Rehberi (TÜBER) [Turkish Nutrition Guidelines]. 2022. Republic of Türkiye Ministry of Health, General Directorate of Public Health, Ministry of Health Publication No: 1031, Ankara. Available from:s URL:https://hsgm.saglik.gov.tr/depo/birimler/saglikli-beslenme-ve-hareketli-hayat-db/Dokumanlar/Rehberler/Turkiye_Beslenme_Rehber_TUBER_2022_min.pdf. Accessed July 18, 2025. Turkish.
21. Guigoz Y, Vellas B, Garry PJ. Assessing the nutritional status of the elderly: The Mini Nutritional Assessment as part of the geriatric evaluation. Nutr Rev. 1996;54:59-65.
22. Sarıkaya D. A Validity Study of Long and Short (MNA-SF) Forms of Mini Nutritional Assessment (MNA) test (thesis). Geriatric Patients Internal Diseasesspecialty. Ankara: Hacettepe University; 2013.
23. Kondrup J, Rasmussen HH, Hamberg O, Stanga Z; Ad Hoc ESPEN Working Group. Nutritional risk screening (NRS 2002): a new method based on an analysis of controlled clinical trials. Clin Nutr. 2003;22:321-36.
24. Bolayır B. A Validity and Reliability Study of Nutritional Assessment Method NRS-2002 (Nutritional Screening Test-2002) in Hospitalized Patients (thesis). Ankara: Hacettepe University; 2014.
25. O'Shea E, Trawley S, Manning E, Barrett A, Browne V, Timmons S. Malnutrition in hospitalised older adults: a multicentre observational study of prevalence, associations and outcomes. J Nutr Health Aging. 2017;21:830-6.
26. Ostrowska J, Sulz I, Tarantino S, Hiesmayr M, Szostak-Węgierek D. Hospital malnutrition, nutritional risk factors, and elements of nutritional care in Europe: comparison of polish results with all European countries participating in the nday survey. Nutrients. 2021;13:263.
27. Mirbazegh S F, Hadavi S, Alipour S, et al. Prevalence of malnutrition among hospitalized elderly patients in hospitals affiliated to Tehran University of Medical Sciences, Tehran, Iran. J Nutr Food Secur. 2022;8:77-82.
28. Bhasin A, Huang L, Shieh MS, Pekow P, Lindenauer PK, Lagu T. Malnutrition in hospitalized adults in the United States, 2016-2019. J Hosp Med. 2024;19:1113-21.
29. Nazlı A, Yılmaz Akyüz E. Assessment of nutritional status of hospitalised geriatric patients and its relationship with sarcopenia. İstanb Med J. 2020;1:290-6.
30. Francis SL, MacNab L, Shelley M. A theory-based newsletter nutrition education program reduces nutritional risk and improves dietary intake for congregate meal participants. J Nutr Gerontol Geriatr. 2014;33:91-107.
31. MacNab L, Francis SL, Lofgren I, et al. Factors influencing dietary intake frequencies and nutritional risk among community-residing older adults. J Nutr Gerontol Geriatr. 2018;37:255-68.
32. Bailey RL, Miller PE, Mitchell DC, et al. Dietary screening tool identifies nutritional risk in older adults. Am J Clin Nutr. 2009;90:177-83.
33. Bellanti F, Lo Buglio A, Quiete S, et al. Comparison of three nutritional screening tools with the new glim criteria for malnutrition and association with sarcopenia in hospitalized older patients. J Clin Med. 2020;9:1898.
34. Mercan Başpınar M, Güden A, Çalık M, Maviş O, Kamat S, Başar H. Determination of malnutrition, frailty, polypharmacy, multimorbidity, and mortality risk in hospitalized elderly COVID-19 pneumonia cases with and without possible sarcopenia. Turk J Geriatr. 2024;27:239-47.
35. Fávaro-Moreira NC, Krausch-Hofmann S, Matthys C, et al. Risk factors for malnutrition in older adults: a systematic review of the literature based on longitudinal data. Adv Nutr. 2016;7:507-22.
36. Bauduer F, Scribans C, Dubernet E, Capdupuy C. Evaluation of the nutritional status of patients over 60-year admitted in a hematology department using the mininutritional assessment

(MNA). A single centre study of 120 cases. J Nutr Health Aging. 2003;7:179-82.

37. Akpınar B, Aydın Cil M, Tosun Taşar P, Karasahin O, Tan G. Frequency of malnutrition and associated risk factors in geriatric patients in a palliative care unit. Palliat Med. 2024;16:48-50.
38. Rashid T, Zia S, Mughal S, Baloch AA, Abdul Rauf MU, Hasan SM. Prevalence of malnutrition and associated factors among the elderly with type 2 diabetes using MNA form. J Nutr Metab. 2025;2025:2107146.
39. Niu M, Zhang F, Wang L, Yang H, Zhu L, Song S. Association of malnutrition risk evaluated by the geriatric nutritional risk index with post-stroke myocardial injury among older patients with firstever ischemic stroke. BMC Geriatr. 2025;25:140.
40. Mengi Çelik Ö, Ağbal Z, Ekici EM. Evaluation of the relationship between malnutrition (MNA-SF), food insecurity, and compliance with the mediterranean diet in community-dwelling older adults. BMC Geriatr. 2025;25:684.
41. Volkert D, Beck AM, Cederholm T, et al. ESPEN guideline on clinical nutrition and hydration in geriatrics. Clin Nutr. 2019;38:10-47.
42. LoBuono DL. The Nutrition and Physical Assessment of Patients with Neurological Disorders (thesis). Kingston; University of Rhode Island; 2014.
43. Demirel B, Atasoy BM. Comparison of three nutritional screening tools to predict malnutrition risk and detect distinctions between tools in cancer patients receiving radiochemotherapy. Nutr Cancer. 2018;70:867-73.

Original Article

DOI: 10.4274/haseki.galenos.2026.30602
Med Bull Haseki 2026;64(4):254-263

Clinical and Radiological Outcomes of Percutaneous K-wire Fixation versus Open Reduction and Internal Fixation for Lisfranc Injuries: A Retrospective Comparative Study

Emre Kocazeybek, Mehmet Ersin

University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Orthopaedics and Traumatology, Istanbul, Türkiye

Abstract

Aim: The optimal fixation strategy for Lisfranc injuries remains controversial, particularly regarding the choice between Kirschner wire (K-wire) fixation and screw-based open reduction and internal fixation (ORIF). This study aimed to compare the clinical, radiological, and complication-related outcomes of closed reduction and percutaneous pinning (CRPP) and ORIF in the treatment of Lisfranc injuries.

Methods: This retrospective comparative study included patients with Lisfranc injuries who were treated with either CRPP using K-wires or ORIF using screws. Demographic characteristics, injury patterns, and operative details were recorded. Functional outcomes were assessed using the American Orthopaedic Foot and Ankle Society (AOFAS) midfoot score, the foot and ankle ability measure-activities of daily living (FAAM-ADL), and the maryland foot score. Radiological evaluation included measurements of intermetatarsal and intercuneiform alignment, dorsal step-off, and tarsometatarsal joint congruity. Postoperative complications and the need for secondary arthrodesis were also analyzed.

Results: A total of 37 patients were included (CRPP: n=17; ORIF: n=20). Patients treated with ORIF demonstrated significantly higher functional scores, including the AOFAS, FAAM-ADL, and Maryland scores, compared with those in the CRPP group (p=0.001). Radiological assessment revealed better maintenance of alignment and a significantly lower dorsal step-off in the ORIF group. Overall complication rates did not differ significantly between groups. However, secondary arthrodesis was more frequently required following CRPP, whereas implant-related events and sensory disturbances tended to occur more often after ORIF, without reaching statistical significance.

Conclusion: Open reduction and internal fixation provided superior functional outcomes and better maintenance of midfoot alignment compared with CRPP. While K-wire fixation may be considered when ORIF is not feasible, rigid fixation with ORIF appears to offer a more reliable restoration of midfoot stability.

Keywords: Lisfranc joint, Kirschner wires, screw fixation, tarsometatarsal joint, foot

Introduction

The Lisfranc joint, also known as the tarsometatarsal (TMT) joint, represents the articulation between the bases of the five metatarsals and the distal surfaces of the three cuneiforms and the cuboid (1). The Lisfranc joint is stabilized by several ligaments, including the dorsal and plantar TMT ligaments and the intermetatarsal ligaments between the second and fifth metatarsals. Lisfranc injuries represent a relatively uncommon but clinically important form of midfoot trauma. Because of the critical role of the TMT joint complex in maintaining midfoot stability and load transmission during gait, disruption of this structure may lead to significant functional impairment (2). Delayed

Corresponding Author: Emre Kocazeybek, MD, University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Orthopaedics and Traumatology, Istanbul, Türkiye
E-mail: emrekcz@gmail.com **ORCID:** orcid.org/0000-0001-5861-8653
Received: 13.03.2026 **Accepted:** 22.06.2026 **Publication Date:** 24.09.2026

Cite this article as: Kocazeybek E, Ersin M. Clinical and radiological outcomes of percutaneous K-wire fixation versus open reduction and internal fixation for Lisfranc injuries: a retrospective comparative study. Med Bull Haseki. 2026;64(4):254-263

diagnosis or inadequate treatment has been associated with chronic midfoot pain, instability, collapse of the longitudinal arch, and post-traumatic osteoarthritis (2,3). Consequently, Lisfranc injuries are considered serious injuries that can result in substantial morbidity and long-term functional limitations if not appropriately managed.

Lisfranc joint injuries often demonstrate instability requiring operative stabilization. While restoration of anatomical alignment is essential for optimal recovery, post-traumatic arthritis continues to be a prevalent postoperative concern (4,5). Various treatment strategies have been proposed for Lisfranc joint injuries, including conservative management for stable injury patterns and surgical techniques such as percutaneous Kirschner wire (K-wire) fixation, open reduction and internal fixation (ORIF), or primary arthrodesis, which is particularly indicated in ligamentous injuries or when the articular surfaces are not reconstructible (6,7).

Open reduction and internal fixation is a well-established technique in the treatment of Lisfranc injuries. However, the need for extensive soft-tissue dissection and the associated risk of soft-tissue complications have driven the development of percutaneous fixation methods (8). In contrast, closed reduction and percutaneous pinning (CRPP) fixation eliminates the need for soft-tissue dissection and offers advantages, such as technical simplicity and a shorter operative time. Nevertheless, this technique has inherent limitations, including inferior biomechanical stability compared with ORIF (9,10).

Although ORIF is widely regarded as the standard surgical treatment for displaced Lisfranc injuries because it allows direct visualization and anatomical reduction, CRPP remains an attractive alternative in selected patients owing to its minimally invasive nature and reduced soft-tissue disruption. However, concerns persist regarding its ability, compared with that of more rigid fixation constructs, to achieve and maintain anatomical reduction. Given the distinct advantages and limitations of these two treatment strategies, comparative evidence regarding their clinical and radiological outcomes remains limited. Accordingly, this study aimed to systematically compare ORIF and CRPP in Lisfranc injuries by evaluating both clinical and radiological parameters, thereby providing a clearer understanding of their relative effectiveness in clinical practice. We hypothesized that CRPP would provide clinical and radiological outcomes comparable to those of screw-based fixation in the treatment of Lisfranc injuries.

Materials and Methods

Compliance with Ethical Standards

The study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval was obtained from the University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital Non-Interventional Clinical Research Ethics Committee (approval no: 9-2026, date: 21.01.2026).

Patients and Study Design

This retrospective cohort study included patients with Lisfranc injuries who underwent CRPP or ORIF between January 2016 and December 2024, as identified from hospital records. The STrengthening the Reporting of observational studies in Epidemiology guidelines were followed when reporting study results (11). Demographic information, clinical data, injury classification, and radiological findings were obtained from the hospital database, patient charts, medical records, operation notes taken during follow-up visits, and radiological X-rays, which were stored in the Picture Archiving and Communication System. The inclusion criteria for this study were as follows: (1) patients diagnosed with Lisfranc injuries treated with either CRPP or ORIF with screws; (2) patients with ideal preoperative and postoperative weight-bearing anteroposterior (AP), oblique, and lateral X-rays. The exclusion criteria were (1) patients treated with plate fixation, suture button fixation, or a combination of fixation methods; (2) patients with open fractures or fracture-dislocations; (3) patients with a history of trauma or prior surgery on the affected foot; (4) inability to undergo radiologic examinations for measurements; and (5) discontinuous follow-ups. During the study period, 82 patients underwent surgery for a Lisfranc injury. As this was a retrospective cohort study, no a priori sample size calculation was performed. Instead, all eligible patients who met the predefined inclusion criteria and had complete clinical and radiological follow-up data were included in the final analysis. Of these, 37 patients met the inclusion criteria and were included in the final analysis.

Exclusions were due to failure to meet the predefined inclusion criteria, including treatment with a plate-screw combination (n=33), suture button fixation (n=3), and inadequate or suboptimal radiographs (n=4). Additionally, patients with discontinuous follow-up data (n=3) and those presenting with open fractures (n=2) were excluded (Figure 1).

The primary outcome of the study was the American Orthopaedic Foot and Ankle Society (AOFAS) Midfoot Score assessed at the final follow-up. Secondary outcomes included the foot and ankle ability measure-activities of daily living (FAAM-ADL) score, the maryland foot score (MFS), radiographic parameters (M1-M2 distance, C1-M2 alignment, C3-M3 alignment, and dorsal step-off), postoperative complications, and the need for secondary arthrodesis.

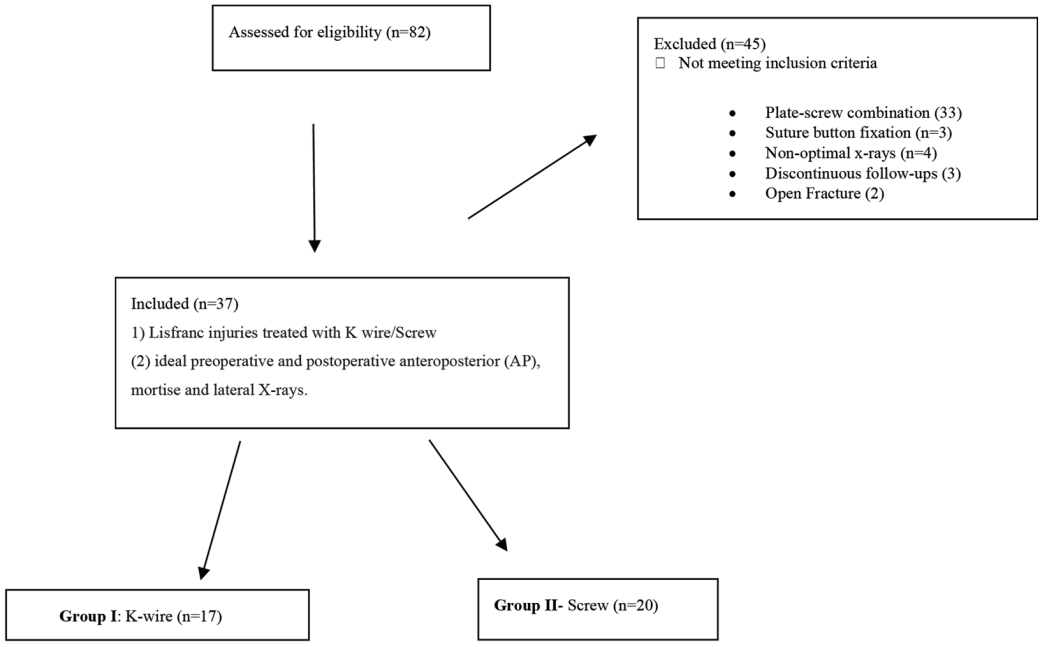

Figure 1. Flow diagram of the study

Surgical Technique

The choice of fixation technique was determined by the treating surgeon according to the fracture characteristics, individual patient factors, and routine clinical practice. Owing to the retrospective nature of the study, no predefined treatment allocation protocol was applied. In the ORIF group, all interventions were performed under general or regional anesthesia and with a tourniquet in the supine position. In cases where the third TMT joint is involved or where instability is present in the fourth or fifth metatarsals, a dual-incision technique is used. This includes a longitudinal incision over the first intermetatarsal space, complemented by a second longitudinal incision aligned with the fourth metatarsal. Reconstruction was performed in a stepwise manner. Intercuneiform instability, when present, was first stabilized with a 3.5-mm cannulated screw. The second metatarsal-intermediate cuneiform articulation, considered the keystone of the TMT complex, was then anatomically reduced using a reduction clamp and temporarily stabilized with K-wires. Definitive fixation was achieved with a 3.5-mm cannulated positioning screw. The Lisfranc articulation was then secured with a 3.5-mm cannulated transarticular screw inserted from the medial cuneiform to the base of the second metatarsal. The first TMT joint was subsequently reduced with a reduction clamp and fixed with one or two 3.5- or 4.5-mm cannulated screws. The remaining TMT joints were addressed sequentially from medial to lateral. The third TMT joint was stabilized when required. Injuries involving the fourth or fifth metatarsals were treated with K-wires inserted from the metatarsal bases into the cuboid, which were typically removed at 6-8 weeks postoperatively (Figure 2).

In the CRPP group, the TMT joints were reduced from medial to lateral, similar to the sequence used in the ORIF group. Manual traction was applied to the distal foot, and the first TMT joint was typically reduced using adduction and plantarflexion maneuvers. After reduction was achieved, a reduction clamp was applied when necessary, and K-wires were inserted percutaneously across the joint to maintain reduction under fluoroscopic guidance (Figure 3). If instability extended to the fourth and fifth TMT joints, these joints were also stabilized with K-wires. Reduction and K-wire positioning were confirmed under fluoroscopic guidance in the AP, lateral, and oblique views. All K-wires were removed at approximately eight weeks postoperatively.

Postoperative Rehabilitation

Postoperative management, including rehabilitation protocols, was individualized according to fracture morphology, fixation stability, and patient-specific factors. Immobilization with a short-leg cast or splint was maintained for 4 weeks. After removal of immobilization, controlled range-of-motion exercises were initiated. Partial weight-bearing was allowed between weeks 6 and 8 postoperatively, with progression to full weight-bearing by 12 weeks.

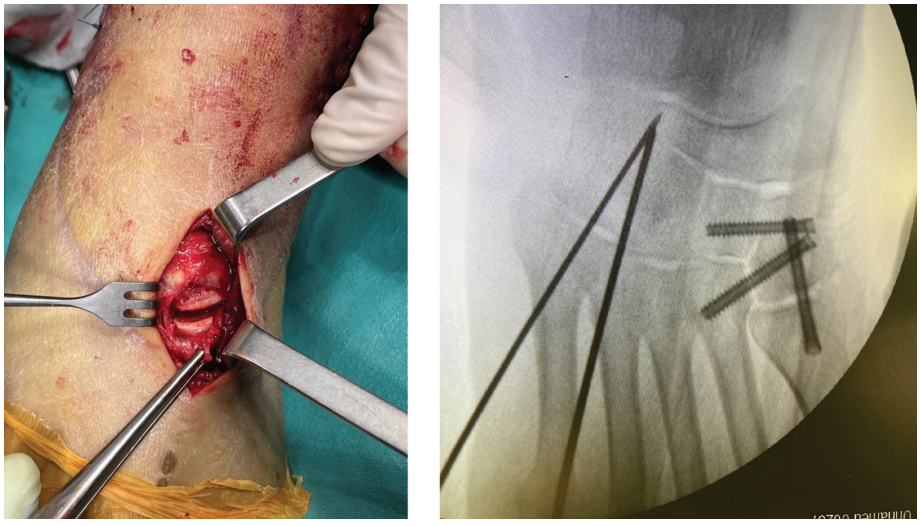

Figure 2. Intraoperative fluoroscopic image demonstrating fixation after open reduction of a Lisfranc injury. The first and second tarsometatarsal joints were stabilized with screw fixation, while the fourth and fifth metatarsals were stabilized using Kirschner wires

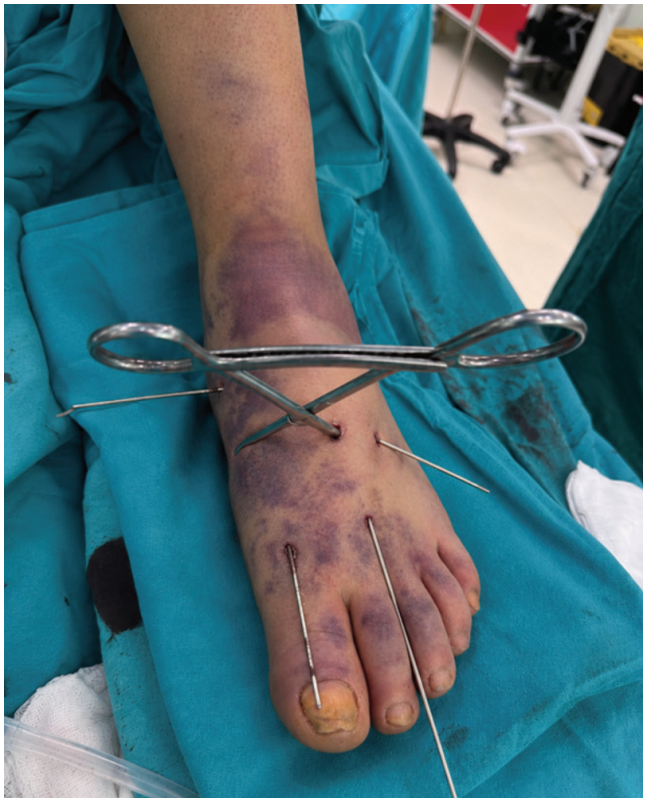

Figure 3. Intraoperative photograph demonstrating closed reduction and percutaneous K-wire fixation of a Lisfranc injury with reduction maintained using a Weber clamp

Radiographic Measurements

Postoperative weight-bearing AP, oblique, and lateral X-rays of the foot were obtained for all patients included in the analysis. Radiological parameters were evaluated in three categories based on AP, oblique, and lateral radiographs.

The first radiological parameter was the diastasis between the first and second metatarsals (M1M2), assessed on weight-bearing AP foot radiographs (Figure 4a-c). According to the instability classification described by Nunley and Vertullo (12), a first-second metatarsal distance of less than 1-2 mm is considered normal, whereas a diastasis of 2-5 mm indicates a subtle injury. A gap of ≥5 mm is typically associated with a reduction in medial arch height and is indicative of a more severe injury. In our study, the intermetatarsal distance was measured on the AP radiograph during the patient's final follow-up, using the criteria defined by Nunley and Vertullo (12).

The second radiographic parameter evaluated on weightbearing AP radiographs was the alignment between the medial cortex of the second metatarsal base and the corresponding medial cuneiform articulation (C1M2) (Figure 5a-c). This relationship was evaluated on the final follow-up weight-bearing AP image and classified as preserved or disrupted based on congruity at the second TMT joint.

The third parameter was assessed on the 30-degree oblique foot radiograph and was defined as the alignment between the lateral cuneiform and the base of the third metatarsal (C3M3) (Figure 6a and b). Specifically, congruity between the medial cortices of the third metatarsal base and the lateral cuneiform was evaluated. Any loss of cortical continuity was interpreted as a disruption of the corresponding TMT articulation. Alignment on the final oblique follow-up radiograph was categorized as preserved or disrupted based on visual assessment of cortical congruity. The radiological parameters assessed

on lateral foot radiographs include the dorsal step-off. The dorsal step-off was measured at the level of the second TMT joint as the absolute distance between the proximal corner of the second metatarsal and the distal corner of the articulating tarsal bone, as previously described by De Bruijn et al. (13) (Figure 7a and b).

Clinical Evaluation

Patients were invited to attend a final visit to undergo a standardized clinical examination and complete four functional outcome questionnaires. All clinical assessments were conducted at the final follow-up by an experienced foot and ankle surgeon. Clinical outcomes were evaluated using the AOFAS Midfoot Score (14), the FAAM-ADL (15) and the MFS (16).

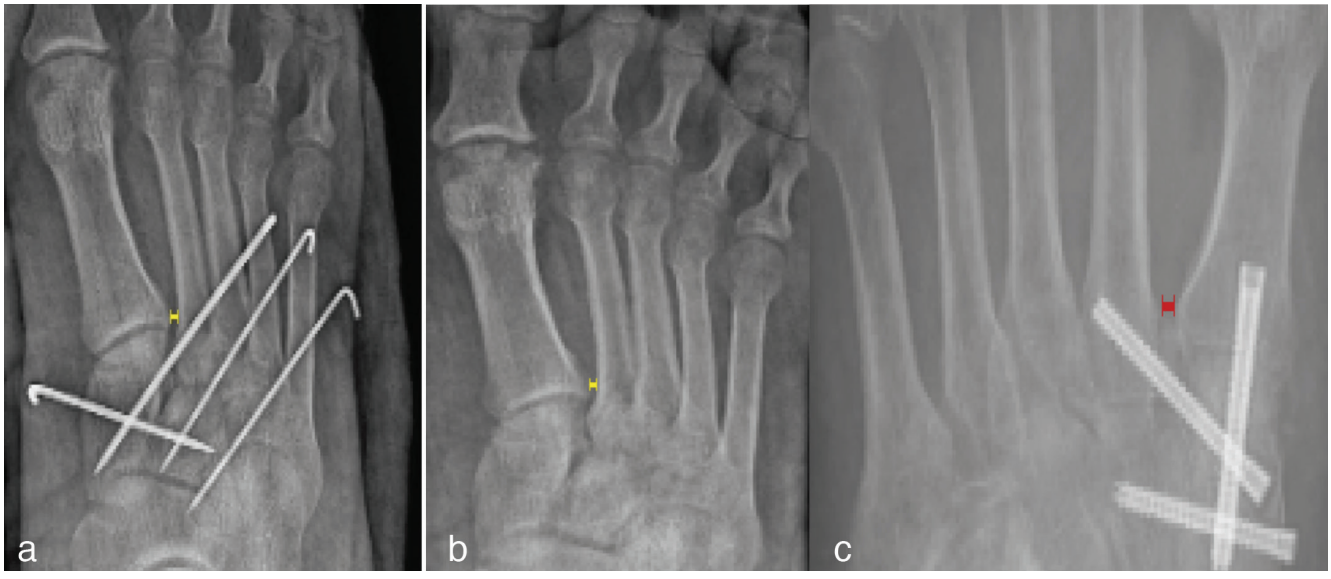

Figure 4. Intermetatarsal distance (M1M2); (a) CRPP group- early postop (yellow line), (b) Late postop following K-wire removal (yellow line), (c) screw group (red line)
CRPP: Closed reduction and percutaneous pinning

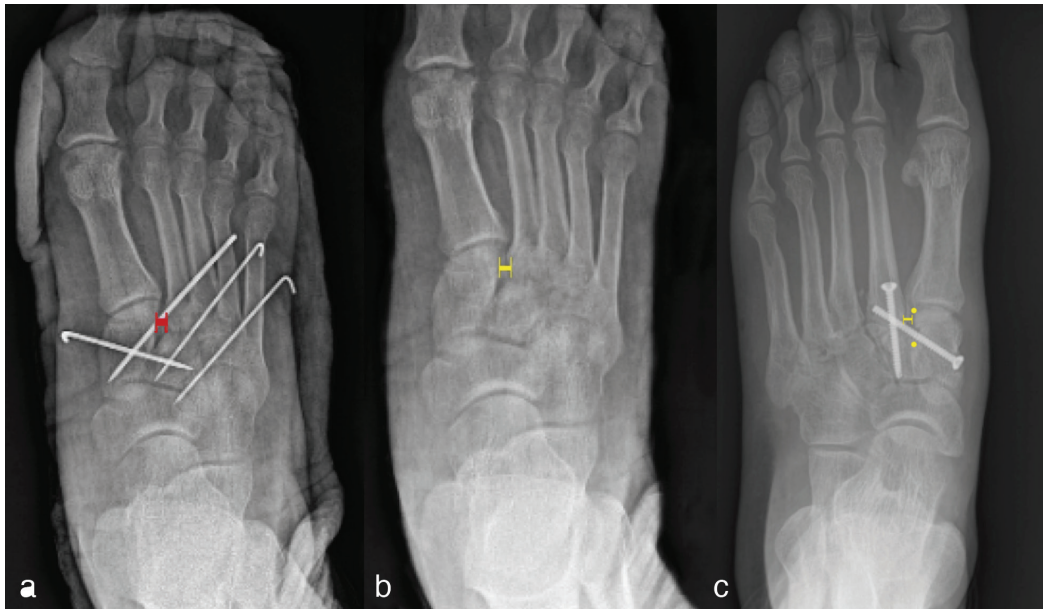

Figure 5. Cuneiform-second metatarsal distance (C1M2); (a) CRPP group- early postop (red line), (b) Late postop following K-wire removal (yellow line), (c) screw group (yellow line)
CRPP: Closed reduction and percutaneous pinning

Statistical Analysis

The study data were summarized using descriptive statistics, including mean, range, and standard deviation for continuous variables and percentages for categorical variables. Statistical analyses were performed using IBM SPSS Statistics version 27.0 (IBM Corp., Armonk, NY, USA). The normality of data distribution was assessed using the Shapiro-Wilk test. Categorical variables were compared between the two groups using the Fisher's exact test. Continuous variables were analyzed using the Student's t-test or the Mann-Whitney U test, as appropriate. All statistical tests were two-tailed, and statistical significance was defined as $p<0.05$.

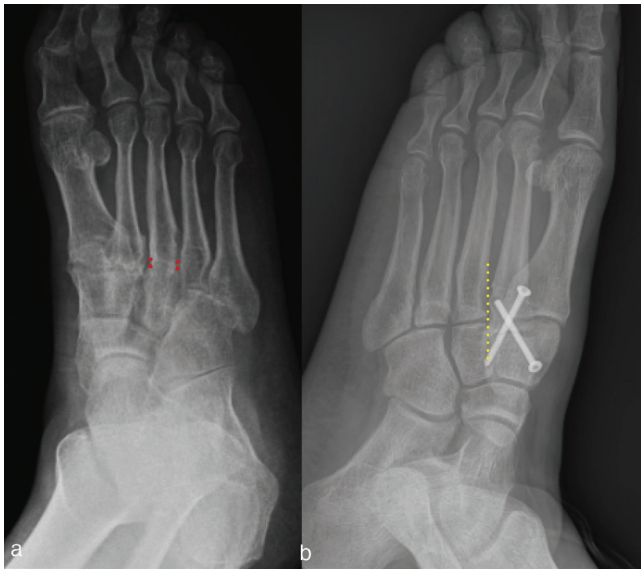

Figure 6. Lateral cuneiform-third metatarsal alignment (C3M3) in oblique X-rays; (a) CRPP group: late postop following K-wire removal (red dots) (b) Screw group (yellow dashed line)
CRPP: Closed reduction and percutaneous pinning

Results

A total of 37 patients were included in the study (CRPP: n=17; ORIF: n=20). No statistically significant differences were observed between the groups with respect to age, sex, body mass index, injury mechanism, Myerson classification, or follow-up duration ($p>0.05$), indicating comparable baseline characteristics (Table 1).

Radiographic evaluation demonstrated significantly better outcomes in the ORIF group, with lower M1-M2 distance and dorsal step-off values than those in the CRPP group ($p=0.012$ and $p=0.006$, respectively). No significant differences were observed in other radiographic parameters (Table 2).

Clinically, the ORIF group showed significantly superior functional outcomes, with higher AOFAS, FAAM-ADL, and Maryland scores ($p\leq0.001$). Ankle dorsiflexion was also significantly greater in the ORIF group ($p=0.013$), whereas plantar flexion did not differ significantly between the groups ($p=0.078$; Table 3).

Complications of CRPP and ORIF Techniques

Overall complication rates were comparable between the groups, with no statistically significant differences observed for individual complications. However, secondary arthrodesis was observed only in the CRPP group (3 cases), a finding that may be clinically relevant despite not reaching statistical significance ($p=0.09$; Table 4).

Discussion

The present study demonstrated that screw-based fixation was associated with superior functional and radiological outcomes compared with CRPP in the treatment of Lisfranc injuries, while overall complication rates were comparable between the two techniques. Patients treated with ORIF showed significantly higher functional scores, including the AOFAS, FAAM-ADL, and Maryland scores, and improved ankle dorsiflexion. Radiologically, the ORIF

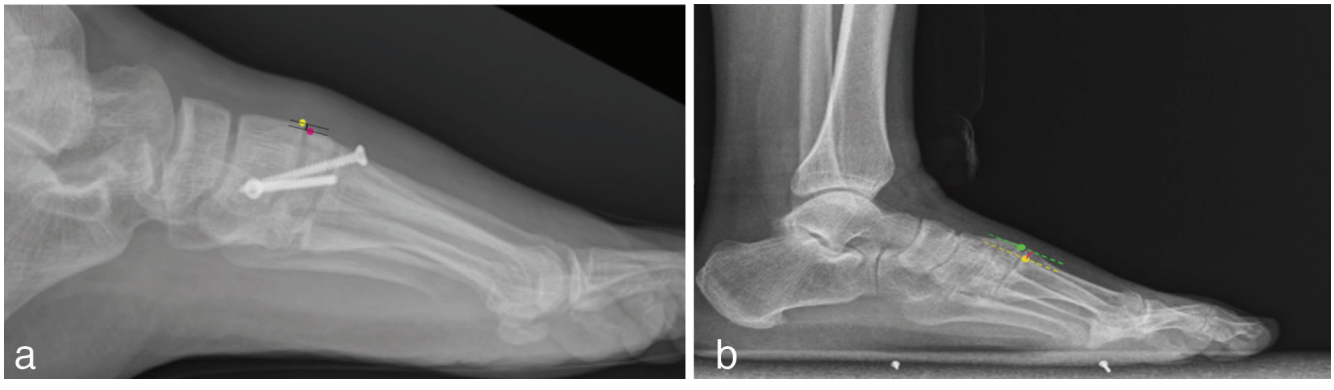

Figure 7. Measurement of the dorsal step-off: Measured as the absolute distance between the proximal corner of the metatarsal bone & the distal corner of the articulating tarsal bone; (a) screw group, (b) CRPP group; late postop following K-wire removal
CRPP: Closed reduction and percutaneous pinning

Table 1. Demographic characteristics of the patients				
		Group 1 CRPP (n=17)	**Group 2 ORIF (n=20)**	**p-value**
Age (year)	**Mean ± SD** **Min-max**	32.57±11.66 (23-61)	36.80±10.90 (20-62)	p=0.625*
Side	**Right (R)** **Left (L)**	9 (52.9%) 8 (47.1%)	10 (50%) 10 (50%)	p=0.88**
Sex	**Female** **Male**	6 (35.2%) 11 (64.8%)	7 (35%) 13 (65%)	p=0.201**
BMI (kg/m²)	**Mean ± SD** **Min-max**	26.8±3.77 (21-34.6)	27.1±3.52 (20.8-35.9)	p=0.358*
Myerson classification **A** **B1** **B2** **C1** **C2**		5 (29.4%) 2 (11.8%) 8 (47.1%) 1 (5.9%) 1 (5.9%)	3 (15%) 1 (5%) 12 (60%) 3 (15%) 1 (5%)	p=0.66**
Injury type **Fall** **Traffic accident** **Crush**		7 (41.2%) 8 (47.1%) 2 (11.8%)	13 (65%) 5 (25%) 2 (10%)	p=0.357**
Follow up (mo)	**Mean ± SD** **Min-max**	26.12±8.13 (12-54)	28.73±7.45 (14-60)	p=0.332*

*Mann-Whitney U **Fisher's exact test
CRPP: Closed reduction and percutaneous pinning, ORIF: Open reduction and internal fixation, BMI: Body mass index, SD: Standard deviation

Table 2. Comparison of radiographic parameters between groups				
		Group 1 CRPP (n=17)	**Group 2 ORIF (n=20)**	**p-value**
M1M2 distance (mm)	**Mean ± SD** **Min-max**	2.02±0.32 (1.1-3.4)	1.65±0.61 (0.7-2.4)	**p=0.012***
C1M2 distance (mm)	**Mean ± SD** **Min-max**	2.01±0.87 (1.2-4.2)	1.57±0.95 (0.6-3.54)	**p=0.15***
C3M3 alignment	**Preserved alignment** **loss of alignment**	13 4	17 3	**p=0.68****
Dorsal step-off (mm)	**Mean ± SD** **Min-max**	1.54±0.47 (0-3.7)	0.62±1.03 (0-1.8)	p=0.006*

*Mann-Whitney U, **Fisher's exact test
CRPP: Closed reduction and percutaneous pinning, ORIF: Open reduction and internal fixation, M1M2: First-second metatarsal distance, C1M2: Medial cortical distance between the medial cuneiform and second metatarsal, C3M3: Medial border alignment of the lateral cuneiform and third metatarsal, SD: Standard deviation

Table 3. Comparison of clinical outcome scores between the CRPP and ORIF groups					
		Group 1 (CRPP) (n=17)	**Group 2 ORIF (n=20)**	**p-value**	**Cohen's d**
AOFAS-midfoot	**Mean ± SD** **Min-max**	70.47±8.25 (61-88)	79.54±7.78 (58-94)	**p=0.001***	1.13
FAAM ADL	**Mean ± SD** **Min-max**	68.47±9.77 (51-88)	78.15±8.73 (58-93)	**p=0.001***	1.05
MFS	**Mean ± SD** **Min-max**	68.63±7.65 (36.18-77.25)	82.96±8.25 (29.57-98.80)	**p=0.001***	1.80
Dorsiflexion	**Mean ± SD** **Min-max**	15.94±3.56 (10-22)	18.55±2.68 (10-25)	**p=0.013***	0.84
PlantarFlexion	**Mean ± SD** **Min-max**	28.31±3.72 (20-35)	30.25±2.78 (20-40)	p=0.078*	0.60

*Mann-Whitney U
CRPP: Closed reduction and percutaneous pinning, ORIF: Open reduction and internal fixation, AOFAS: American Orthopaedic Foot and Ankle Society, FAAM-ADL: Foot and ankle ability measure-activities of daily living, SD: Standard deviation, MFS: Maryland foot score

Table 4. Comparison of postoperative complications between the CRPP and ORIF groups			
	Group 1 CRPP (n=17)	**Group 2 ORIF (n=20)**	**p-value**
Implant breakage (n)	0	2	0.49**
Compartment syndrome (n)	0	0	NA
Reflex sympathetic dystrophy (n)	4	2	0.38**
Superficial infection (n)	0	3	0.79**
Deep infection (n)	0	1	1.00**
Malunion (n)	4	1	0.15**
Reoperation-secondary arthrodesis (n)	3	0	0.09**

**Fisher's Exact test
CRPP: Closed reduction and percutaneous pinning, ORIF: Open reduction and internal fixation

group demonstrated significantly lower M1-M2 distances and dorsal step-off values, indicating better maintenance of anatomical reduction. Contrary to our initial hypothesis, CRPP did not demonstrate clinical and radiological outcomes comparable to those of screw-based fixation. Although CRPP offers the theoretical advantages of reduced soft-tissue disruption and a less invasive surgical approach, the present findings suggest that these benefits may not fully compensate for the reduced mechanical stability associated with K-wire fixation. In summary, these results suggest that more rigid fixation constructs may be beneficial, especially in adult patients with unstable Lisfranc injuries.

Functional outcome scores are key indicators of recovery after Lisfranc injuries and closely correlate with the quality of anatomical reduction. In this context, several studies have evaluated the clinical outcomes of different fixation techniques. Sánchez-Gómez et al. (17) reported significantly higher AOFAS scores in patients treated with screw fixation compared to those managed with K-wires, suggesting that improved stability and maintenance of reduction may contribute to better functional outcomes. Similarly, Uzun et al. (18) demonstrated that screw fixation provided more favorable radiological alignment and maintenance of reduction, which may be associated with improved clinical results over time. Conversely, Dudko et al. (19) reported satisfactory functional outcomes following K-wire fixation, particularly in low-energy injuries, emphasizing that acceptable results can be achieved when anatomical reduction is properly obtained. In addition, a multicenter study by Adachi et al. (20) found comparable functional outcomes between K-wire and screw fixation techniques, indicating that neither method showed clear superiority in terms of functional recovery. In contrast to adult populations, a study in pediatric and adolescent patients reported favorable midterm outcomes regardless of fixation method, with K-wires being commonly used and sufficient when anatomical reduction is achieved (21). These findings suggest that fixation type may be less critical in younger populations, where biological healing potential and lower mechanical demands may play a greater role in recovery. In the present study, however, screw-based fixation was associated with significantly superior clinical outcomes across multiple functional scoring systems, including AOFAS, FAAM-ADL, and Maryland scores, compared with percutaneous K-wire fixation. This finding suggests that in adult patients, more rigid fixation may be necessary to achieve and maintain optimal alignment and to facilitate improved functional recovery in Lisfranc injuries.

Radiological evaluation is essential for assessing the quality of reduction and long-term stability of the Lisfranc joint complex, with parameters such as intermetatarsal distance, joint alignment, and dorsal step-off, which closely correlate with clinical outcomes. In this context, Ghate et al. (22) reported that both screw and K-wire fixation could achieve satisfactory results; however, they emphasized that the accuracy and maintenance of anatomical reduction were the primary determinants of successful outcomes rather than the fixation method alone. A recent study by Talia et al. (23) highlighted that while both K-wire fixation and transarticular screw fixation are commonly used, screw-based constructs tend to provide greater stability and may be more effective in maintaining anatomical reduction over time, particularly in unstable injury patterns. The authors also emphasized that loss of reduction remains a key factor associated with inferior outcomes, further emphasizing the necessity of stable fixation. The findings of this study indicate that ORIF was associated with improved radiological alignment, as reflected by lower M1-M2 distances and dorsal step-off values. These results suggest that more rigid fixation may facilitate more reliable restoration and maintenance of anatomical reduction, which may ultimately contribute to improved functional outcomes.

Postoperative complications represent an important consideration when comparing fixation techniques for Lisfranc injuries. Previous studies indicate that Kirschner-wire fixation and screw fixation have largely comparable complication profiles, although certain differences exist.

For instance, screw fixation has been associated with a higher rate of implant-related complications, whereas K-wire fixation tends to be associated with minor complications, particularly in low-energy injuries with adequate reduction (17,19). Multicenter studies have suggested that complications such as malunion, infection, and secondary arthrodesis may occur regardless of fixation method, highlighting the influence of injury severity and reduction quality. More recent evidence further indicates that postoperative complications are more strongly associated with patient- and injury-related factors than with the fixation technique itself (20,24). In line with these findings, overall complication rates were similar between the CRPP and ORIF groups in this study. After ORIF, implant-related problems and sensory disturbances were more common. After CRPP, alignment-related problems and secondary arthrodesis were more common. However, none of these differences were statistically significant, and the infection rates were similar between the two groups.

The higher incidence of secondary arthrodesis after CRPP may be clinically meaningful despite the lack of statistical significance, as secondary arthrodesis is generally regarded as an indicator of persistent instability, post-traumatic degeneration, or inadequate maintenance of midfoot alignment following Lisfranc injury treatment (23). In the present study, secondary arthrodesis was required in three patients in the CRPP group, whereas no patients in the ORIF group required secondary arthrodesis during follow-up. Although ORIF demonstrated a lower rate of secondary arthrodesis in our cohort, recent meta-analyses have shown that secondary arthrodesis may still be necessary in a subset of patients following ORIF despite satisfactory initial reduction and fixation, likely due to the development of post-traumatic degeneration over time (23,25,26). Supporting this observation, Almaat et al. (27) recently evaluated 8,101 patients who underwent ORIF for Lisfranc injuries and reported a remarkably low long-term conversion rate to secondary arthrodesis (2.7% at 5 years and 2.8% at 10 years). Despite a substantial rate of implant-related reoperations, the low incidence of secondary fusion suggests that joint-preserving ORIF provides durable midfoot stability over time for the majority of patients.

Study Limitations

This study has several limitations that should be considered when interpreting its findings. First, the retrospective design and the absence of randomization may have introduced selection bias and limited the ability to establish causal relationships. In addition, a substantial proportion of initially screened patients were excluded based on predefined eligibility criteria, primarily because they underwent alternative fixation methods or lacked adequate radiographic follow-up. Although these criteria were applied to create a more homogeneous study population, the relatively high exclusion rate may have contributed to selection bias and limited the generalizability of the findings.

Second, operations were performed by different surgeons during the study period, and variations in surgical technique and decision-making may have influenced outcomes. Third, a single foot-and-ankle specialist who was not blinded to the study conducted the clinical evaluations, which may have introduced assessment bias. Furthermore, radiographic measurements were not assessed for interobserver or intraobserver reliability, which may affect the reproducibility of these findings because all measurements were performed by a single observer. Fourth, the relatively small sample size may have resulted in an underpowered analysis, limiting the ability to detect statistically significant differences, particularly in less frequent outcomes such as complications and secondary arthrodesis. Additionally, the small sample size precluded the use of multivariable analysis, which may have limited the ability to fully account for potential confounding factors. Although functional outcomes were evaluated using established scoring systems, the minimal clinically important difference was not formally assessed, which should be considered when interpreting the clinical relevance of the results. Despite these limitations, this study provides a direct comparison of two commonly used fixation strategies—open reduction with screw fixation and percutaneous K-wire fixation—and contributes meaningful clinical and radiological data to the literature. Although the sample size is relatively small, the findings underscore the potential importance of rigid fixation in achieving improved outcomes for patients with Lisfranc injuries.

Conclusion

Our findings suggest that screw-based fixation is associated with superior functional outcomes and better maintenance of midfoot alignment compared with K-wire fixation. Although K-wire fixation may be considered an alternative in selected patients, particularly when open reduction cannot be performed, ORIF was associated with improved outcomes in this cohort. However, future adequately powered multicenter prospective studies are warranted.

Ethics

Ethics Committee Approval: Ethical approval was obtained from the University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital Non-Interventional Clinical Research Ethics Committee (approval no: 9-2026, date: 21.01.2026).

Informed Consent: Written informed consent for the use of anonymized clinical data was obtained from all patients.

Footnotes

Authorship Contributions

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

Conflict of Interest: The authors declared that there were no conflicts of interest.

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

References

1. Moracia-Ochagavía I, Rodríguez-Merchán EC. Lisfranc fracture-dislocations: current management. EFORT Open Rev. 2019;4:430-44.
2. McDermott A, Repanshek Z, Koyfman A, Long B. High risk and low incidence diseases: Lisfranc injury. Am J Emerg Med. 2024;85:172-8.
3. Poutoglidou F, van Groningen B, McMenemy L, Elliot R, Marsland D. Acute Lisfranc injury management. Bone Joint J. 2024;106-B:1431-42.
4. Sinkler MA, Benedick A, Kavanagh M, Vallier HA. Functional outcomes after high-energy lisfranc injuries. Foot Ankle Int. 2023;44:960-7.
5. Smith N, Stone C, Furey A. Does open reduction and internal fixation versus primary arthrodesis improve patient outcomes for Lisfranc trauma? A systematic review and meta-analysis. Clin Orthop Relat Res. 2016;474:1445-52.
6. Hammad A, Ahmad Y, Abdelnour J. Lisfranc injuries: latest updates on diagnostics and management. Transl Sports Med. 2026;2026:3933956.
7. Balboni JM, Levine AR, Boggiano VJ, et al. Operative treatment of Lisfranc injuries in elite athletes: 2024 International Foot and Ankle Sports Consensus and systematic review. Knee Surg Sports Traumatol Arthrosc. 2026;34:3089-105.
8. Perugia D, Basile A, Battaglia A, Stopponi M, De Simeonibus AU. Fracture dislocations of Lisfranc's joint treated with closed reduction and percutaneous fixation. Int Orthop. 2003;27:30-5.
9. Sinkler MA, Benedick A, Kavanagh M, Alfonso N, Vallier HA. Complications and outcomes after fixation of Lisfranc injuries at an urban level 1 trauma center. J Orthop Trauma. 2024;38:e169-74.
10. Lee CA, Birkedal JP, Dickerson EA, Vieta PA Jr, Webb LX, Teasdall RD. Stabilization of Lisfranc joint injuries: a biomechanical study. Foot Ankle Int. 2004;25:365-70.
11. Vandenbroucke JP, von Elm E, Altman DG, et al.; STROBE Initiative. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. Epidemiology. 2007;18:805-35.
12. Nunley JA, Vertullo CJ. Classification, investigation, and management of midfoot sprains: Lisfranc injuries in the athlete. Am J Sports Med. 2002;30:871-8.
13. De Bruijn J, Hagemeijer NC, Rikken QGH, et al. Lisfranc injury: refined diagnostic methodology using weightbearing and non-weightbearing radiographs. Injury. 2022;53:2318-25.
14. Kitaoka HB, Alexander IJ, Adelaar RS, Nunley JA, Myerson MS, Sanders M. Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. Foot Ankle Int. 1994;15:349-53.
15. Martin RL, Irrgang JJ, Burdett RG, Conti SF, Van Swearingen JM. Evidence of validity for the foot and ankle ability measure (FAAM). Foot Ankle Int. 2005;26:968-83.
16. Maryland foot score. J Orthop Trauma. 2006;20:S96-7.
17. Sánchez-Gómez P, Lajara-Marco F, Salinas-Gilabert JE, Lozano-Requena JA. Lisfranc fracture-dislocation: screw vs. K-wire fixation. Rev Esp Cir Ortop Traumatol. 2008;52:130-6.
18. Uzun E, Misir A, Kizkapan TB, Ozcamdalli M, Arat F. Evaluation of mid-term clinical and radiologic outcomes after open reduction and internal fixation of Lisfranc fracture-dislocations: a case series. Ann Med Res. 2019;26:1615-21.
19. Dudko S, Kusz D, Pierzchaa A. Lisfranc injury-fixation with Kirschner wires. Foot and Ankle Surgery. 2004;10:5-8.
20. Adachi A, Takegami Y, Nakashima H, Mishima K, Kobayashi K, Imagama S. Comparative clinical outcomes of K-wire fixation versus screw fixation in Lisfranc joint injuries: a multicenter (TRON group) retrospective study. J Foot Ankle Surg. 2025;64:683-7.
21. Sensoz E, Yilmaz H, Onay T. Mid-term radiologic and clinical Results of pediatric-adolescent lisfranc injuries. J Pediatr Orthop. 2025;45:e405-12.
22. Ghate SD, Sistla VM, Nemade V, Vibhute D, Shahane SM, Samant AD. Screw and wire fixation for Lisfranc fracture dislocations. J Orthop Surg (Hong Kong). 2012;20:170-5.
23. Talia AJ, Roshan-Zamir S, Shepherd DW. Outcomes in open reduction and internal fixation of tarsometatarsal (Lisfranc) injuries: a contemporary review of fixation constructs. Foot Ankle Orthop. 2025;10:24730114251352519.
24. Minoura Y, Takegami Y, Nakashima H, Mishima K, Mabuchi M, Imagama S. Risk factors for postoperative complications in Lisfranc injuries treated with open reduction and internal fixation: a multicenter study. J Foot Ankle Surg. 2026;65:159.e1-5.
25. Levy CJ, Yatsonsky D 2nd, Moral MZ, Liu J, Ebraheim NA. Arthrodesis or open reduction internal fixation for Lisfranc injuries: a meta-analysis. Foot Ankle Spec. 2022;15:179-84.
26. Vesely BD, Michels L, King MA, Gangopadhyay P, Scott AT. Primary arthrodesis versus open reduction and internal fixation for lisfranc joint fracture-dislocations in adults: a systematic review. J Am Podiatr Med Assoc. 2024; 114:21232.
27. Almaat A, Jawich B, Cavataio J, et al. Do Lisfranc ORIFs fail? Low long-term conversion to arthrodesis in the largest ORIF-specific cohort to date. Injury. 2026;57:113187.

Original Article

DOI: 10.4274/haseki.galenos.2026.93063
Med Bull Haseki 2026;64(4):264-270

Diagnostic Performance of Neutrophil-to-lymphocyte, Platelet-to-lymphocyte, and Monocyte-to-lymphocyte Ratios in Aseptic Loosening after Hip and Knee Arthroplasty

Bunyamin Ari[1], Erhan Akdemir[2], Ahmed Arif Uzun[3], Arda Bilir[1]

[1]*Kutahya Health Sciences University Faculty of Medicine, Department of Orthopedics and Traumatology, Kutahya, Türkiye*
[2]*Kutahya Health Sciences University Faculty of Medicine, Department of Nuclear Medicine, Kutahya, Türkiye*
[3]*Kutahya City Hospital, Clinic of Orthopaedics and Traumatology, Kutahya, Türkiye*

Abstract

Aim: Aseptic loosening remains one of the leading causes of late failure after hip and knee arthroplasty, and its diagnosis is often challenging because imaging and conventional inflammatory markers have limited specificity. We aimed to evaluate and compare the diagnostic performance of neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and monocyte-to-lymphocyte ratio (MLR) in detecting aseptic loosening following hip and knee arthroplasty.

Methods: This single-center, retrospective diagnostic accuracy study included 291 patients who underwent three-phase bone scintigraphy for suspected prosthetic loosening between 2020 and 2025. Patients were classified into aseptic loosening and control groups based on scintigraphic findings supported by clinical evaluation. Hematological parameters were obtained from complete blood counts, and NLR, PLR, and MLR were calculated. Diagnostic performance was assessed using receiver operating characteristic curve analysis, and optimal cut-off values were determined using the Youden index.

Results: In knee arthroplasty, MLR was significantly higher and the white blood cell count was significantly lower in patients with aseptic loosening, whereas NLR and PLR showed no significant differences. In hip arthroplasty, both PLR and MLR were significantly elevated in the loosening group. Receiver operating characteristic analysis demonstrated the low-to-moderate discriminatory ability of MLR in both knee (AUC=0.61) and hip (AUC=0.64) arthroplasty, whereas PLR achieved the highest diagnostic performance in the hip group (AUC=0.68) with high sensitivity. Neutrophil-to-lymphocyte ratio consistently demonstrated limited diagnostic value across both joint types.

Conclusion: Monocyte-to-lymphocyte ratio and PLR demonstrate modest diagnostic utility in the evaluation of aseptic loosening after hip and knee arthroplasty, with joint-specific differences in performance. These hematological markers should be considered adjunctive tools rather than standalone diagnostic tests and may contribute to clinical decision-making when interpreted alongside imaging findings and clinical assessment.

Keywords: Aseptic loosening, prosthesis failure, monocyte-to-lymphocyte ratio, platelet-to-lymphocyte ratio, biomarkers, radionuclide imaging

Introduction

Degenerative joint diseases represent a major and growing public health burden worldwide, largely driven by population aging and increased life expectancy. Osteoarthritis, the most prevalent form, is a leading cause of chronic pain, disability, and loss of functional independence in individuals over 60 years of age (1-3). Total knee arthroplasty (TKA) and total hip arthroplasty

Corresponding Author: Lec, Bunyamin Ari, MD, Kutahya Health Sciences University Faculty of Medicine, Department of Orthopedics and Traumatology, Kutahya, Türkiye
E-mail: bunyamin.ari@ksbu.edu.tr **ORCID:** orcid.org/0000-0001-7638-7932
Received: 29.10.2025 **Accepted:** 14.03.2026 **Publication Date:** 24.09.2026

Cite this article as: Ari B, Akdemir E, Uzun AA, Bilir A. Diagnostic performance of neutrophil-to-lymphocyte, platelet-to-lymphocyte, and monocyte-to-lymphocyte ratios in aseptic loosening after hip and knee arthroplasty. Med Bull Haseki. 2026;64(4):264-270

(THA) are well-established and highly effective surgical treatments for end-stage joint degeneration, providing substantial improvements in pain relief and quality of life (4). However, the increasing volume of primary arthroplasty procedures has been accompanied by a parallel rise in revision surgeries, among which aseptic loosening remains one of the most common late failure mechanisms (5).

Accurate diagnosis of aseptic loosening is clinically challenging. Conventional diagnostic approaches rely on a combination of clinical assessment and imaging modalities, including plain radiographs, three-phase bone scintigraphy (TPBS), single-photon emission computed tomography combined with computed tomography (SPECT/CT), and magnetic resonance imaging (6). Although these techniques are widely used, they may be limited by cost, availability, metal-related artifacts, and suboptimal sensitivity for early loosening (7). Consequently, there is growing interest in simple, minimally invasive, and cost-effective biomarkers that could support clinical decision-making. In this context, inflammatory indices derived from routine complete blood count parameters—such as the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and monocyte-to-lymphocyte ratio (MLR)—have attracted attention in orthopedic research. While these markers have been extensively investigated in the differentiation of periprosthetic joint infection from aseptic failure, their diagnostic value specifically in non-infectious aseptic loosening remains insufficiently defined, with inconsistent and heterogeneous results across studies (8).

Most previous investigations have focused on either hip or knee arthroplasty populations in isolation and have rarely compared multiple hematological markers within the same methodological framework. We hypothesized that inflammatory cell ratios, particularly MLR, differ between patients with and without aseptic loosening and that they demonstrate joint-specific diagnostic performance for hip and knee arthroplasty. This study aimed to compare the diagnostic accuracy of NLR, PLR, and MLR in patients with suspected aseptic loosening following TKA and THA using a uniform diagnostic approach, thereby clarifying their relative clinical utility across joint types and supporting more informed adjunctive use of readily available hematological parameters in the diagnostic evaluation of prosthetic loosening.

Materials and Methods

Compliance with Ethical Standards

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Due to its retrospective design and the use of anonymized data obtained from routine clinical records, formal written informed consent was waived. Ethical approval was granted by the Kutahya Health Sciences University Non-Interventional Clinical Research Ethics Committee (approval no.: 2024/14-11, date: 16.12.2024). No additional diagnostic or therapeutic procedures were performed for research purposes.

Study Design

This single-center retrospective diagnostic accuracy study evaluated hematological inflammatory ratios for identifying aseptic loosening after total hip and knee arthroplasty. The study included patients who underwent TPBS for suspected prosthetic loosening between January 2020 and May 2025. The patient selection process and study flow are summarized in Figure 1.

Study Population and Patient Selection

All eligible consecutive patients who underwent TPBS for suspected prosthetic loosening during the study period were screened to minimize selection bias. We excluded 21 patients because of incomplete records or diagnoses other than aseptic loosening, resulting in a final cohort of 291 patients.

Group 1 (loosening): One hundred and eleven (73 TKA, 38 THA) with scintigraphic findings suggestive of aseptic loosening confirmed by clinical evaluation. Clinical evaluation included assessment of prosthesis stability, pain score, and plain radiographs; when ambiguity persisted, SPECT/CT was performed using a Siemens Symbia Intevo scanner to obtain reference uptake indices.

Group 2 (control): One hundred and eighty (112 TKA, 68 THA) with TPBS performed for suspected prosthetic

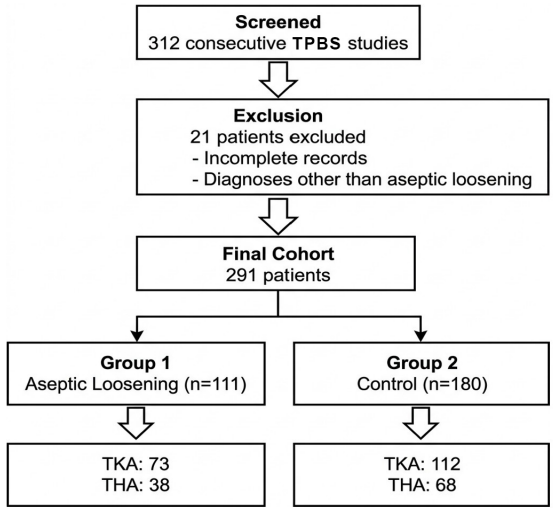

Figure 1. Study flow diagram illustrating patient selection and group allocation
TKA: Total knee arthroplasty, THA: Total hip arthroplasty, TPBS: Three-phase bone scintigraphy

loosening but reported as "normal" or "not in favor of loosening." Because the number of eligible control patients exceeded that of the loosening group, controls were selected by simple random sampling from an anonymized list using a computer-generated random number table to minimize selection bias.

Imaging Protocol

In this study, TPBS was performed in patients evaluated for suspected loosening after total joint arthroplasty. Scintigraphic imaging was performed using a Mediso gamma camera system. Prior to imaging, patients were intravenously administered 20-25 mCi (740-925 MBq) of technetium-99m labeled methylene diphosphonate. The TPBS protocol included a dynamic blood flow phase during the first minute post-injection, followed by a blood pool phase at 3-5 minutes, and delayed static images acquired approximately 3 hours after injection. In interpreting the scans, the absence of increased activity or hyperemia in the blood flow and pool phases was considered indicative of the absence of infection. However, in the delayed phase, increased tracer uptake in the periprosthetic bone compared with adjacent healthy bone, particularly in areas under mechanical stress, was interpreted as suggestive of aseptic loosening. Based on this imaging protocol, the final diagnosis of aseptic loosening was established by integrating scintigraphic findings with clinical and laboratory data.

Diagnostic Definition of Aseptic Loosening

Aseptic loosening was diagnosed based on increased periprosthetic tracer uptake during the delayed phase of TPBS in the absence of clinical, laboratory, or imaging findings suggestive of infection. Importantly, histopathological confirmation and intraoperative cultures were not routinely available, and the diagnosis relied on the integration of imaging and clinical data. The absence of histopathological confirmation may have led to misclassification bias, a methodological limitation of the study.

Data Collection and Laboratory Measurements

Demographic variables (age, sex), arthroplasty type (hip or knee), and time interval between primary surgery and scintigraphic evaluation were recorded. Laboratory parameters, including white blood cell, neutrophil, lymphocyte, monocyte, and platelet counts, and C-reactive protein (CRP) levels, were retrieved from the hospital information system. To minimize temporal variability, laboratory measurements obtained closest to the scintigraphic examination (within ±7 days) were included in the analysis. When multiple measurements were available, the value closest to the imaging date was used.

Derived inflammatory ratios were calculated as follows:

- NLR = Neutrophil-to-lymphocyte ratio
- PLR = Platelet-to-lymphocyte ratio
- MLR = Monocyte-to-lymphocyte ratio

Outcome Measures

The primary outcome was the diagnostic performance of NLR, PLR, and MLR in distinguishing patients with aseptic loosening from controls, assessed using receiver operating characteristic (ROC) curve analysis.

The secondary outcomes included:

- Comparison of hematological parameters between loosening and control groups stratified by joint type (hip vs. knee),
- Determination of optimal cut-off values using the Youden index,
- Evaluation of sensitivity and specificity for each marker.

Statistical Analysis

Statistical analyses were performed using the SPSS, version 25.0 (IBM Inc., Armonk, NY, USA). Normality of continuous variables was assessed using the Kolmogorov-Smirnov test. Normally distributed variables were expressed as mean ± standard deviation and compared using independent samples t-tests. Categorical variables were compared using the chi-square test.

Receiver operating characteristic curve analysis was used to assess diagnostic accuracy, and area under the curve (AUC) values were calculated with 95% confidence intervals (CIs). Optimal cut-off values were determined using the Youden index.

Given the exploratory nature of this study and the evaluation of multiple biomarkers and subgroup analyses, no formal correction for multiple comparisons was applied; the results should be interpreted accordingly. A p-value <0.05 was considered statistically significant.

Results

Demographic and Clinical Characteristics

A total of 291 patients were included in the analysis, comprising 111 patients with aseptic loosening and 180 controls. In the loosening group, 73 patients had TKA and 38 patients had THA. Baseline demographic characteristics, including age, sex distribution, and the time interval from primary surgery to scintigraphic evaluation, were comparable between the loosening and control groups for both joint types (Table 1).

Hematological Findings in Knee Arthroplasty

In the TKA cohort, patients with aseptic loosening demonstrated significantly higher MLR values compared with controls (0.26±0.12 vs. 0.21±0.06, p=0.024). The white blood cell count was significantly lower in the loosening group (9.34±3.67 vs. 14.45±4.87 $\times 10^9$/L, p=0.031). No significant differences were observed for neutrophil count, lymphocyte count, platelet count, NLR, or PLR (Table 2).

Receiver operating characteristic analysis showed that MLR had low-to-moderate discriminative ability for detecting aseptic loosening in the knee cohort (AUC=0.61, 95% CI: 0.53-0.69). NLR and PLR demonstrated poor diagnostic performance (AUC <0.60) (Figure 2).

Hematological Findings in Hip Arthroplasty

In the THA cohort, both PLR and MLR were significantly higher in patients with aseptic loosening compared with controls (PLR: p=0.033; MLR: p=0.017). No significant differences were observed in white blood cell count, neutrophil count, lymphocyte count, platelet count, NLR, or CRP levels (Table 2).

Receiver operating characteristic analysis revealed that PLR had the highest diagnostic accuracy in the hip cohort (AUC=0.68, 95% CI: 0.58-0.78), followed by MLR (AUC=0.64, 95% CI: 0.53-0.75). NLR again showed limited discriminative value (AUC=0.56; Figure 2).

Table 1. Demographic and clinical characteristics of patients with suspected aseptic loosening following hip and knee arthroplasty

Group (n=291)	Number of men (n=159)	Number of women (n=132)	Age (Mean ± SD)	Follow-up period (Mean ± SD years)
Knee patient (73)	41	32	68.3±12.7	11.3±7.2
Knee control (112)	61	51	70.6±9.8	10.2±8.3
Hip patient (38)	21	17	65.4±16.2	13.1±8.9
Hip control (68)	36	32	69±11.6	12.1±4.1

Values are presented as mean ± standard deviation or number of patients
SD: Standard deviation

Table 2. Comparison of demographic, clinical, and hematological parameters between patients with and without aseptic loosening after knee and hip arthroplasty

	Variables	Group 1		Group 2	p
Knee	Gender	Male	41	61	0.825
		Sex/female	32	51	
	Age		68.3±12.7	70.6±9.8	0.642
	Follow-up period (year)		11.3±7.2	10.2±8.3	0.562
	Hematological parameters	NLR	2.35±1.71	1.93±0.84	0.282
		PLR	141.12±70.37	136.00±81.65	0.167
		MLR	0.26±0.12	0.21±0.06	**0.024**
		CRP	7.3±5.4	9.8±10.8	0.072
		White blood cell	9.34±3.67	14.45±4.87	**0.031**
Hip	Gender	Male	21	36	0.218
		Sex/female	17	32	
	Age		65.4±16.2	69.0±11.6	0.432
	Follow-up period (year)		13.1±8.9	12.1±4.1	0.136
	Hematological and biochemistry parameters	NLR	2.04±1.27	2.43±1.69	0.07
		PLR	143.42±13.58	131.00±28.42	**0.033**
		MLR	0.22±0.19	0.14±0.09	**0.017**
		CRP	9.6±6.1	8.5±4.8	0.231
		White blood cell	10.71±3.92	11.44±6.72	0.855

Values are presented as mean ± standard deviation or number of patients. Group 1: patients with aseptic loosening; Group 2: controls. Comparisons between groups were performed using the independent-samples t-test for continuous variables and the chi-square test for categorical variables. Bold p-values indicate statistically significant differences (p<0.05). Age is given in years, CRP in mg/L, and white blood cell count in $\times 10^9$/L
NLR: Neutrophil-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, MLR: Monocyte-to-lymphocyte ratio, CRP: C-reactive protein

Optimal Cut-off Values

Optimal cut-off values derived using the Youden index are summarized in Table 3. In knee arthroplasty, an MLR threshold of 0.268 yielded a sensitivity of 43.8% and a specificity of 86.6%. In hip arthroplasty, a PLR cut-off of 122.85 provided high sensitivity (97.4%) with low specificity (44.1%), whereas an MLR cut-off of 0.278 demonstrated high specificity (94.1%) with moderate sensitivity (55.3%).

Subgroup and Sensitivity Analyses

Subgroup analyses stratified by sex and follow-up duration (>10 years vs. ≤10 years) did not reveal significant interactions affecting MLR or PLR performance (p for interaction >0.1). Sensitivity analyses excluding patients with CRP levels >10 mg/L yielded comparable AUC values, supporting the robustness of the findings.

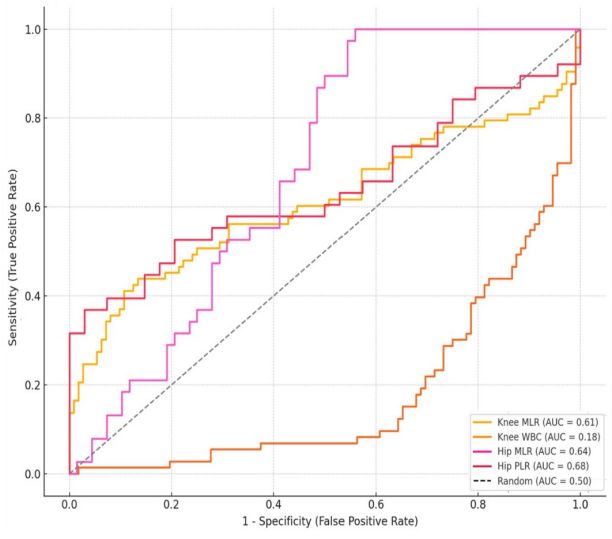

Figure 2. Receiver operating characteristic curves demonstrating the diagnostic performance of hematological parameters in the detection of aseptic loosening following hip and knee arthroplasty

MLR: Monocyte-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, *AUC: Area under the curve, WBC: White blood cell*

Discussion

In this single-center retrospective study, we evaluated the diagnostic performance of routinely available hematological inflammatory ratios in patients with suspected aseptic loosening following total hip and knee arthroplasty. The main finding of this study is that the MLR demonstrated modest but consistent diagnostic performance in both joint types, whereas the NLR showed limited value in the non-infectious setting. In contrast, the PLR exhibited joint-specific behavior, showing clinically relevant performance only in hip arthroplasty (9,10).

The relatively better performance of MLR may be explained by the underlying pathophysiology of aseptic loosening. Chronic low-grade inflammation induced by wear debris leads to macrophage activation and periprosthetic osteolysis, processes in which monocytes play a central role (7,11). As monocytes are precursors of macrophages, MLR may better reflect these biological mechanisms compared with NLR, which is more closely associated with acute infectious responses (12). This distinction likely explains why NLR, despite its established utility in periprosthetic joint infection, failed to demonstrate diagnostic relevance in the present aseptic cohort (13,14).

An important finding of this study is the differential diagnostic performance of PLR in hip versus knee arthroplasty. While PLR did not show significant value in knee arthroplasty, it demonstrated high sensitivity in hip arthroplasty, suggesting a potential role as a rule-out marker in this subgroup (15,16). Differences in joint biomechanics, implant characteristics, and local inflammatory responses may contribute to this finding, although the exact mechanism remains unclear. Given its low specificity, PLR should not be used as a standalone diagnostic parameter but may provide supportive information when combined with more specific markers such as MLR.

Although several hematological parameters reached statistical significance, their clinical implications should be interpreted with caution. The values of the area under the ROC curve for MLR and PLR ranged from 0.61 to 0.68, indicating fair but limited discriminatory ability. These findings suggest that hematological ratios alone are insufficient for definitive diagnosis and should be interpreted as adjunctive tools alongside clinical assessment and imaging modalities (17,18).

Table 3. Optimal cut-off values, sensitivity, and specificity of inflammatory markers for the diagnosis of aseptic loosening based on Youden index analysis

Parameter	Cut-off value	Sensitivity	Specificity
Knee MLR	0.268	0.438	0.866
Hip PLR	122.85	0.974	0.441
Hip MLR	0.278	0.553	0.941

Cut-off values were derived from receiver operating characteristic curve analysis. Sensitivity and specificity values reflect the diagnostic performance of each parameter in knee and hip arthroplasty cohorts
MLR: Monocyte-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio

Previous studies have primarily focused on differentiating periprosthetic joint infection from aseptic failure or have evaluated hip and knee arthroplasty populations separately (19,20). By contrast, the present study simultaneously analyzed both joint types using a uniform diagnostic methodology and directly compared multiple inflammatory ratios within the same cohort. This approach provides a more comprehensive assessment of the potential role of hematological markers in the diagnostic work-up of aseptic loosening.

Study Limitations

This study has several limitations. Its retrospective, single-center design may limit generalizability and introduce selection bias. Aseptic loosening was diagnosed based on clinical assessment and TPBS, without histopathological confirmation or intraoperative cultures; this may have resulted in misclassification bias. In addition, multiple biomarkers were analyzed without formal adjustment for multiple comparisons, and multivariate modeling was not performed.

Despite these limitations, the study benefits from a relatively large cohort, the inclusion of both hip and knee arthroplasty patients, and the evaluation of simple, widely available hematological markers using a standardized diagnostic approach.

Conclusion

Monocyte-to-lymphocyte ratio and PLR showed modest diagnostic value for assessing aseptic loosening after hip and knee arthroplasty, with joint-specific differences in performance. MLR demonstrated higher specificity across both joint types, whereas PLR showed high sensitivity, mainly in hip arthroplasty. Given the moderate discriminatory ability of these markers (AUC: 0.61-0.68), they should be considered adjunctive tools rather than standalone diagnostic tests. Future prospective multicenter studies with surgical or histopathological confirmation are warranted to validate these findings.

Ethics

Ethics Committee Approval: Ethical approval was granted by the Kutahya Health Sciences University Non-Interventional Clinical Research Ethics Committee (approval no.: 2024/14-11, date: 16.12.2024).

Informed Consent: Since the study was designed retrospectively, individual informed consent was waived.

Footnotes

Authorship Contributions

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

Conflict of Interest: The authors declared that there were no conflicts of interest.

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

References

1. Ricciardi G, Siracusano L, Micale E, et al. Aseptic loosening in total hip arthroplasty: pathophysiology, biomarkers, and preventive treatment strategies. Applied Sciences. 2025;15:9156.
2. Yu BZ, Li R, Li X, Chai W, Zhou YG, Chen JY. The relationship of C-reactive protein/interleukin-6 concentrations between serum and synovial fluid in the diagnosis of periprosthetic joint infection. J Orthop Surg Res. 2021;16:733.
3. Köksal A, Öner A, Çimen O, et al. Femoral stem fractures after primary and revision hip replacements: a single-center experience. Jt Dis Relat Surg. 2020;31:557-63.
4. Moldovan F. Sterile inflammatory response and surgery-related trauma in elderly patients with subtrochanteric fractures. Biomedicines. 2024;12:354.
5. Zhao G, Chen J, Wang J, et al. Predictive values of the postoperative neutrophil-to-lymphocyte ratio, platelet-to-lymphocyte ratio, and lymphocyte-to-monocyte ratio for the diagnosis of early periprosthetic joint infections: a preliminary study. J Orthop Surg Res. 2020;15:571.
6. Buijs GS, Kooijenga AC, Rikken QGH, Schafroth MU, Kievit AJ, Blankevoort L. MRI and SPECT/CT demonstrate, with low certainty of evidence, the highest diagnostic accuracy for aseptic knee arthroplasty loosening: a systematic comparative diagnostic test review and meta-analysis. Knee Surg Sports Traumatol Arthrosc. 2024;32:2061-74.
7. Larikka MJ, Ahonen AK, Niemelä O, et al. Comparison of 99mTc ciprofloxacin, 99mTc white blood cell and three-phase bone imaging in the diagnosis of hip prosthesis infections: improved diagnostic accuracy with extended imaging time. Nucl Med Commun. 2002;23:655-61.
8. Maimaiti Z, Xu C, Fu J, Chai W, Zhou Y, Chen J. The potential value of monocyte to lymphocyte ratio, platelet to mean platelet volume ratio in the diagnosis of periprosthetic joint infections. Orthop Surg. 2022;14:306-14.
9. Kildow BJ, Springer BD, Brown TS, Lyden ER, Fehring TK, Garvin KL. Long term results of two-stage revision for chronic periprosthetic knee infection: a multicenter study. J Arthroplasty. 2022;37:S327-32.
10. Chalmers PN, Sporer SM, Levine BR. Correlation of aspiration results with periprosthetic sepsis in revision total hip arthroplasty. J Arthroplasty. 2014;29:438-42.
11. Cai C, Zeng W, Wang H, Ren S. Neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR) and monocyte-to-lymphocyte ratio (MLR) as biomarkers in diagnosis evaluation of acute exacerbation of chronic obstructive pulmonary disease: a retrospective, observational study. Int J Chron Obstruct Pulmon Dis. 2024;19:933-43.
12. Tirumala V, Klemt C, Xiong L, Chen W, van den Kieboom J, Kwon YM. Diagnostic utility of platelet count/lymphocyte count ratio and platelet count/mean platelet volume ratio in

periprosthetic joint infection following total knee arthroplasty. J Arthroplasty. 2021;36:291-7.

13. Moldovan F. Role of serum biomarkers in differentiating periprosthetic joint infections from aseptic failures after total hip arthroplasties. J Clin Med. 2024;13:5716.
14. Festa E, Ascione T, Bernasconi A, et al. Diagnostic performance of neutrophil to lymphocyte ratio, monocyte to lymphocyte ratio, platelet to lymphocyte ratio, and platelet to mean platelet volume ratio in periprosthetic hip and knee infections: a systematic review and meta-analysis. Diagnostics (Basel). 2022;12:2033.
15. Deng L, Wang J, Yang GY, et al. Different biomarker ratios in peripheral blood have limited value in diagnosing periprosthetic joint infection after total joint arthroplasty: a single-center, retrospective study. BMC Musculoskelet Disord. 2024;25:377.
16. Tosun MF, Çiçeklidağ M, Bircan R, et al. The role of inflammatory markers in the differential diagnosis of stiff shoulder disease. Jt Dis Relat Surg. 2025;36:415-9.
17. Wier J, Jones IA, Palmer R, et al. Complete blood cell count-based ratios identify total joint arthroplasty patients likely to benefit from perioperative dexamethasone. J Bone Joint Surg Am. 2025;107:163-73.
18. Li J, Shang H, Ma T, et al. Case report: prosthetic revision due to aseptic loosening following total knee arthroplasty: a clinical management and pathological mechanism investigation. Front Surg. 2026;12:1710878.
19. Vílchez-Cavazos F, Rodríguez-Corpus LA, Arrambide-Garza FJ, Acosta-Olivo CA, Simental-Mendía M. Diagnostic accuracy of the monocyte-to-lymphocyte ratio in periprosthetic joint infection: a systematic review and meta-analysis. Arch Orthop Trauma Surg. 2026;146:13.
20. Li GQ, Zhang J, Huang Y. Factors associated with aseptic loosening after primary total hip arthroplasty: a systematic review and meta-analysis. World J Orthop. 2026;17:114482.

Original Article

DOI: 10.4274/haseki.galenos.2026.09326
Med Bull Haseki 2026;64(4):271-284

Upper Extremity Pain, Disability, Sleep Quality, and Professional Quality of Life Among Healthcare Workers Delivering Physical Therapy Modalities: A Cross-sectional Study

Selim Sezikli, Yelda Soluk Ozdemir

University of Health Sciences Türkiye, Istanbul Physical Therapy and Rehabilitation Training and Research Hospital, Clinic of Physical Medicine and Rehabilitation, Istanbul, Türkiye

Abstract

Aim: Healthcare workers delivering physical therapy modalities perform repetitive hand–arm movements under static loading, yet they are usually studied together with physiotherapists. This study aimed to evaluate the relationships among neck and upper extremity pain, upper extremity disability, sleep quality, and professional quality of life in this workforce.

Methods: This single-center, cross-sectional study was conducted between January 2024 and May 2024 at a tertiary training and research hospital; 65 healthcare workers delivering physical therapy modalities were analyzed. Pain, disability, sleep quality, and professional quality of life were assessed with the Numeric Rating Scale, QuickDASH, Pittsburgh Sleep Quality Index, and Professional Quality of Life Scale. Spearman correlation, with false discovery rate correction, and multivariable linear regression were used.

Results: Mean age was 27.8±8.9 years and 80.0% were female. The median pain score was 4 for both regions, and 58.5% had poor sleep quality. Upper extremity pain was correlated with disability (r=0.769) and with sleep quality (r=0.562). Pain, professional experience, sleep quality, and inappropriate posture were independent correlates of disability (adjusted R^2=0.665), whereas pain and shorter sleep duration were independent correlates of sleep quality (adjusted R^2=0.363).

Conclusion: Pain intensity was most consistently associated with both disability and sleep quality, whereas cumulative exposure and posture, rather than the modality applied, distinguished participants. An ergonomic review of the working day, combined with a brief sleep screen, may be warranted for this workforce.

Keywords: Occupational diseases, musculoskeletal diseases, upper extremity, sleep quality, burnout professional, physical therapy modalities

Introduction

Work-related musculoskeletal disorders (WRMSDs) are among the most frequent occupational health problems in healthcare; recent evidence syntheses place their burden at the upper end of the occupational spectrum. A 2025 meta-analysis of 40 studies involving 19,903 nurses reported a pooled prevalence of 84.3% [95% confidence interval (CI) 81.1-87.4], with the neck and shoulder being among the three most affected regions (1). Additionally, a continental meta-analysis of 68 studies involving 17,188 surgeons published in the same year reported pooled prevalences ranging from 62.8% to 77.6% depending on the region (2). Among physiotherapists, a global meta-analysis of 26 studies reported prevalences of 35.4% for the thumb, 26.4% for the neck, 20.8% for the shoulder, and 18.1% for the wrist and hand (3), and a study of physiotherapists in Madrid found that 98% reported at least one musculoskeletal symptom within the previous 12 months (4). Across these syntheses the risk factors are

Corresponding Author: Selim Sezikli, MD, University of Health Sciences Türkiye, Istanbul Physical Therapy and Rehabilitation Training and Research Hospital, Clinic of Physical Medicine and Rehabilitation, Istanbul, Türkiye
E-mail: selimsezikli@hotmail.com **ORCID:** orcid.org/0000-0002-5380-4972
Received: 26.01.2026 **Accepted:** 23.08.2026 **Publication Date:** 24.09.2026

Cite this article as: Sezikli S, Soluk Ozdemir Y. Upper extremity pain, disability, sleep quality, and professional quality of life among healthcare workers delivering physical therapy modalities: a cross-sectional study. Med Bull Haseki. 2026;64(4):271-284

consistent: repetitive hand-arm movements, sustained static and non-neutral postures, insufficient recovery between tasks, and high workload intensity (5,6).

Healthcare workers whose daily work involves delivering physical therapy modalities—ultrasound, transcutaneous electrical nerve stimulation (TENS), extracorporeal shock wave therapy (ESWT), shortwave diathermy, and infrared therapy—have an exposure profile that differs from that of physiotherapists because their working day is composed almost entirely of short, repeated device applications with limited task variation. Despite this distinction, they are almost always pooled with physiotherapists rather than analyzed separately, and data specific to this group remain scarce (4). Upper extremity pain in this setting is unlikely to be an isolated peripheral problem: a systematic review with meta-analysis has established a bidirectional relationship between sleep problems and chronic musculoskeletal pain (6), sleep quality has been proposed as a mediator of work-related musculoskeletal discomfort (7), and a recent pilot study using the Pittsburgh Sleep Quality Index (PSQI) documented clinically relevant sleep impairment in rehabilitation staff (8). To our knowledge, no study has examined whether pain, upper extremity disability, sleep quality, and professional quality of life covary in this workforce, and which of these variables carry independent weight.

We hypothesized that, in this occupational group, higher intensity of upper extremity and neck pain would be associated with greater upper extremity disability and poorer sleep quality, and that psychosocial burden—characterized by higher burnout and compassion fatigue together with lower compassion satisfaction—would be independently associated with both outcomes after adjustment for pain and occupational exposure. The aim of this study was, therefore, to evaluate the relationships among upper extremity and neck pain, upper extremity disability, sleep quality, and professional quality of life, and to identify the independent correlates of disability and sleep quality among healthcare workers delivering physical therapy modalities. By distinguishing the variables that carry independent weight from those that do not, this study may help inform the design of targeted preventive programs in which ergonomic measures and psychosocial support are directed at the factors that matter for this workforce.

Materials and Methods

Compliance with Ethical Standards

This study was approved by the University of Health Sciences Türkiye, Bakirkoy Dr. Sadi Konuk Training and Research Hospital Clinical Research Ethics Committee (approval number: 2023-15-15, date: 07.08.2023). Data were collected at Istanbul Physical Medicine and Rehabilitation Training and Research Hospital. The study was conducted in accordance with the principles of the Declaration of Helsinki as revised in 2013, and written informed consent was obtained from all participants before data collection. All data for this study were collected primarily through self-administered questionnaires; no routinely recorded or retrospectively retrieved clinical data were used. The authors declare no conflict of interest, and the study received no financial support. The study was designed, conducted, and reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement for cross-sectional studies (9).

Study Design

This single-center, cross-sectional, observational study was conducted between January 2024 and May 2024 at the University of Health Sciences Türkiye, Istanbul Physical Medicine and Rehabilitation Training and Research Hospital. The unit operated on a fixed daytime schedule during the study period, and all staff worked day shifts only. The flow of participants from the eligible workforce to the analyzed sample is presented in Figure 1.

Participants

Personnel aged between 18 and 65 years who had been actively delivering physical therapy modalities for at least six months were eligible for participation. Individuals with a history of cancer, rheumatological disease, or neurological disease, those who had experienced major upper extremity trauma or undergone surgery within the previous three months, pregnant individuals, and those using psychiatric medications known to affect sleep were excluded from the study. All eligible personnel working in the unit during the study period were invited consecutively; 65 provided written informed consent and completed all questionnaires. The total number of eligible personnel employed during the study period could not be retrieved retrospectively; consequently, the participation rate could not be calculated, and volunteer bias could not be excluded, as acknowledged in the study limitations. All participants held the same clinical role within the physical therapy unit and applied the same range of therapeutic modalities; they differed only in educational background (two-year versus four-year higher-education programs). There were no missing data, and all analyses were carried out on the full sample of 65 participants.

Data Collection

During data collection, participants completed a structured questionnaire assessing sociodemographic characteristics, including age, sex, marital status, height, weight, smoking and alcohol use, years of professional

experience, weekly working hours, and self-reported sleep duration. Body mass index was calculated from recorded height and weight. Sleep duration was recorded in ordinal bands: 1-4 hours, 5-6 hours, 7-8 hours, or more than 9 hours. For each of the five modalities routinely used in clinical practice—ultrasound, TENS, infrared therapy, shortwave diathermy, and ESWT—use was recorded as a binary variable based on participants' self-reports. Self-reported mean daily application time was recorded only for ultrasound and ESWT using ordinal bands (0-30 min, 30 min-1 h, 1-2 h, 2-3 h, 3-4 h, 4-5 h, or more than 5 h); no duration data were collected for TENS, infrared therapy, or shortwave diathermy. Participants also reported whether they took rest breaks during work, whether they remained in the same position for prolonged periods, and whether they maintained an appropriate body posture while working.

Assessment Tools

Upper extremity function was assessed using the QuickDASH questionnaire, an 11-item instrument evaluating functional difficulties, symptom severity, and the impact of upper extremity symptoms on daily activities (10). Scores range from 0 to 100, with higher scores indicating greater disability. The optional four-item work module was administered in addition to the main scale and was scored on the same 0-100 metric. The Turkish version of the QuickDASH has been shown to be valid and reliable (11).

Sleep quality was measured using the PSQI, which evaluates seven components: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, sleep disturbances, use of sleep medication, and daytime dysfunction (12). Each component is scored from 0 to 3 and the global score ranges from 0 to 21; a global score greater than 5 indicates poor sleep quality. For the global score and for each component, higher scores denote poorer sleep. The Turkish validity and reliability study of the PSQI was conducted by Ağargün et al. (13).

Professional quality of life was assessed using the Turkish version of the Professional Quality of Life Scale, Revision IV (ProQOL R-IV) (14,15). The instrument comprises 30 self-report items that form three independent 10-item subscales—compassion satisfaction, burnout, and compassion fatigue—with each item rated on a six-point frequency scale ranging from 0 (never) to 5 (very often), giving subscale scores between 0 and 50. Items 1, 4, 15, 17, and 29 are reverse scored. The subscales are evaluated separately, and no total score is calculated. Higher compassion satisfaction scores indicate a more positive professional experience, whereas higher burnout and

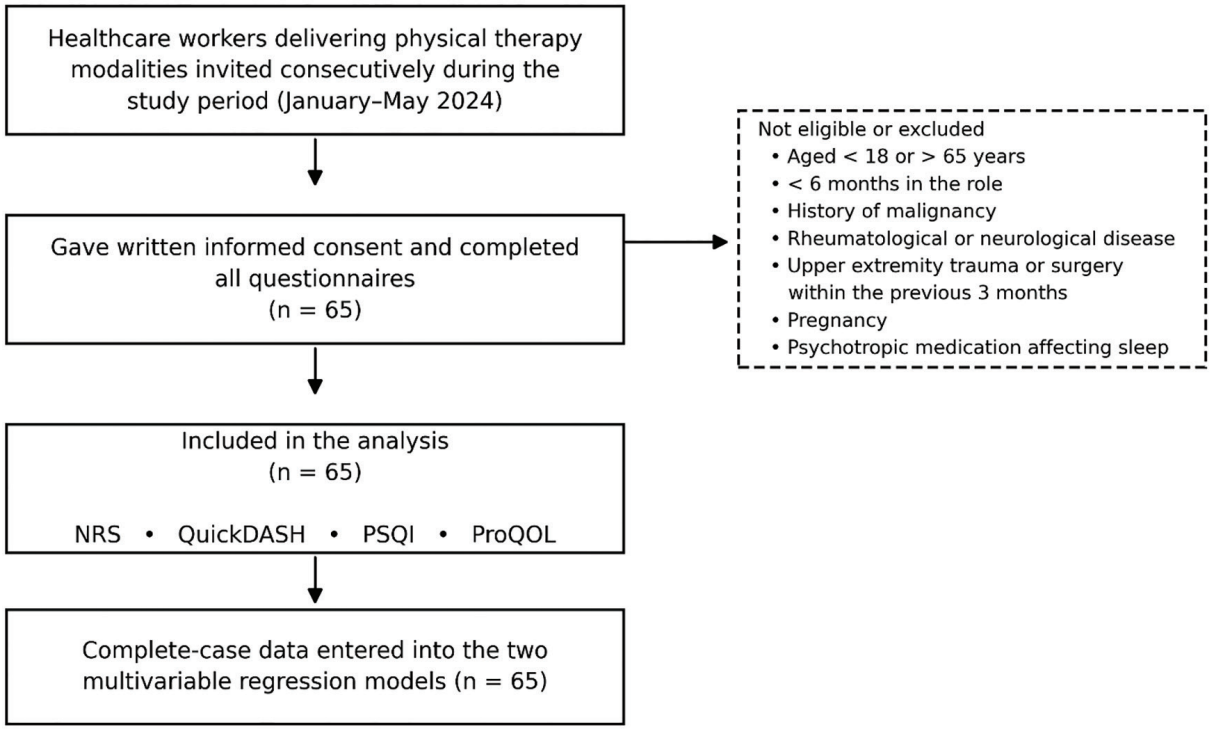

Figure 1. Flow of participants through the study
NRS: Numeric Rating Scale, PSQI: Pittsburgh Sleep Quality Index, ProQOL: Professional Quality of Life Scale, QuickDASH: Quick disabilities of the arm, shoulder and hand questionnaire

compassion fatigue scores indicate greater psychosocial burden. In the Turkish validation study, Cronbach's alpha coefficients were 0.82 for compassion satisfaction, 0.62 for burnout, and 0.84 for compassion fatigue (15).

Pain intensity was assessed using the Numeric Rating Scale (NRS), an 11-point integer scale ranging from 0 to 10, where 0 represented "no pain" and 10 represented "the most severe pain imaginable". Participants rated their mean pain intensity over the previous week separately for the upper extremity and the neck.

Statistical Analysis

Statistical analyses were performed using SPSS version 22.0 (IBM Corp., Armonk, NY, USA). The distribution of continuous variables was assessed using the Shapiro-Wilk test and visual inspection of histograms and Q-Q plots. Normally distributed variables are presented as mean ± standard deviation and non-normally distributed variables as median interquartile range (IQR); categorical variables are presented as frequency and percentage. Group comparisons were conducted using the independent samples t-test or the Mann-Whitney U test, depending on the data distribution. The chi-square test was used for categorical variables. Correlations were evaluated using the Spearman rank test; for dichotomous variables, this is equivalent to the point-biserial coefficient, and coding directions are given in the table footnotes. To control for multiple testing, the Benjamini-Hochberg false discovery rate correction was applied within each correlation table at q=0.05, and both unadjusted p values and adjusted q values are reported (16). Two-sided p-values below 0.05 were considered statistically significant.

Two multivariable linear regression models were constructed using the enter method, with the QuickDASH score and the PSQI global score as dependent variables. Candidate predictors were selected based on clinical relevance and bivariate associations at $p<0.10$; to limit overfitting, the number of predictors was capped at five per model, resulting in 13 participants per predictor. Neck pain and age were not entered because of collinearity with upper extremity pain (rho=0.66) and with years of professional experience (rho=0.85), respectively. Initial analysis showed that the residuals of the QuickDASH model were not normally distributed (Shapiro-Wilk $p<0.001$). A square-root transformation was applied to the dependent variable, correcting the distribution ($p=0.136$). The assumptions of linear regression, including normality of residuals, homoscedasticity, multicollinearity, and independence of observations, were assessed using residual plots and Q-Q plots, the Shapiro-Wilk test, the Durbin-Watson statistic, and variance inflation factor (VIF) and tolerance values. Cook's distance was used to screen for influential observations. Model fit was assessed using the F test, R^2, and adjusted R^2. Regression coefficients are reported as unstandardized B (with 95% CIs), standard errors, standardized β coefficients, t values, and p-values.

Since the sample size was limited to the available workforce at the study center, a post-hoc power analysis was performed using G*Power 3.1 (17). With 65 participants and a two-sided alpha of 0.05, the study had 80% power to detect a correlation of $r\geq0.34$, which corresponds to a medium effect. For a regression model with five predictors, the same sample could detect an effect of $f^2\geq0.22$ ($R^2 \approx 0.18$). Each model included 13 participants per predictor, which is above the conventional minimum of ten. The study was therefore sensitive to moderate and large effects, whereas it may have been underpowered to detect weaker associations.

Results

Demographic and Occupational Characteristics

A total of 65 healthcare workers who delivered physical therapy modalities were analyzed. Their demographic and occupational characteristics are given in Table 1, and the modalities they applied, together with daily application durations, are given in Table 2. The mean age was 27.8±8.9 years (median 26, IQR 23-29; range 18-54). Fifty-two participants (80.0%) were female, and 30 participants (46.2%) had completed a four-year higher education program. The median duration of professional experience was 2 years (IQR 2-6; range 1-36) and the median weekly working time was 40 hours (range 24-45). Thirty-two participants (49.2%) reported sleeping 5-6 hours per night, another 32 (49.2%) reported 7-8 hours, and one participant (1.5%) reported sleeping 1-4 hours per night. Rest breaks during work were reported by 57 participants (87.7%), while 48 (73.8%) said they worked in the same position for prolonged periods, and only 31 (47.7%) said they maintained an appropriate body posture. The median NRS score was 4 for both upper extremity pain (IQR 2-5) and neck pain (IQR 3-6); both had a range of 1-9. Thirty-eight participants (58.5%) had a PSQI global score above 5, indicating poor sleep quality.

Bivariate Associations with Disability and Sleep Quality

Pain intensity was the variable most consistently associated with both outcome domains. Every association reported in this paragraph remained significant after correction for multiple testing (Table 3). Upper extremity pain correlated with the QuickDASH score (r=0.769), the QuickDASH work module score (r=0.764), and the PSQI global score (r=0.562). It was also associated with five PSQI components: subjective sleep quality (r=0.458), use of sleep medication (r=0.434), daytime dysfunction (r=0.433), sleep disturbances (r=0.420), and sleep latency

Table 1. Demographic and occupational characteristics of the participants (n=65)		
Variable	**Category**	**Value**
Age (years)		27.8±8.9; median 26 (23-29); range 18-54
Sex, n (%)	Female	52 (80.0)
	Male	13 (20.0)
Marital status, n (%)	Married	21 (32.3)
	Single	44 (67.7)
Body mass index (kg/m^2)[a]		23.7±3.8
Educational attainment, n (%)	Two-year higher education program	35 (53.8)
	Four-year higher education program	30 (46.2)
Smoking, n (%)	Never	43 (66.2)
	Occasionally	15 (23.1)
	Frequently	7 (10.8)
Alcohol use, n (%)	Never	58 (89.2)
	Occasionally	6 (9.2)
	Frequently	1 (1.5)
Pain intensity (NRS, 0-10), median (IQR)	Upper extremity	4 (2.0-5.0); range 1-9
	Neck	4 (3.0-6.0); range 1-9
Self-reported sleep duration, n (%)	1-4 hours	1 (1.5)
	5-6 hours	32 (49.2)
	7-8 hours	32 (49.2)
Years of professional experience, median (IQR)		2 (2-6); range 1-36
Weekly working hours, median (IQR)		40 (40-40); range 24-45
Rest breaks during work, n (%)	Yes	57 (87.7)
	No	8 (12.3)
Prolonged working in the same position, n (%)	Yes	48 (73.8)
	No	17 (26.2)
Maintaining appropriate body posture, n (%)	Yes	31 (47.7)
	No	34 (52.3)
QuickDASH score, median (IQR)		17.5 (6.0-31.2)
QuickDASH work module score, median (IQR)		12.5 (6.2-31.2)
PSQI global score, median (IQR)		6 (4-9.0)
Poor sleep quality (PSQI >5), n (%)		38 (58.5)
ProQOL compassion satisfaction, median (IQR)		37 (31-44.0)
ProQOL burnout, median (IQR)		14 (9-19.0)
ProQOL compassion fatigue, median (IQR)		10 (5-19.0)

[a]Body mass index was calculated from the recorded height and weight values. Normality was assessed with the Shapiro-Wilk test. Normally distributed variables are presented as mean ± standard deviation, non-normally distributed variables as median (interquartile range), and categorical variables as n (%)
IQR: Interquartile range, NRS: Numeric Rating Scale, ProQOL: Professional Quality of Life Scale, PSQI: Pittsburgh Sleep Quality Index, QuickDASH: Quick disabilities of the arm, shoulder and hand questionnaire

(r=0.325). Neck pain followed the same pattern for the QuickDASH score (r=0.659), the work module (r=0.642), the PSQI global score (r=0.479), subjective sleep quality (r=0.417), sleep disturbances (r=0.423), and daytime dysfunction (r=0.430), but not for sleep latency. Since higher PSQI scores denote poorer sleep, these positive coefficients indicate that greater pain was associated with worse sleep across all affected components.

Among the occupational and behavioral variables, longer professional experience was associated with greater use of sleep medication (r=0.459, q=0.001) and with a higher QuickDASH score (r=0.325, q=0.040). Participants

Table 2. Physical therapy modalities applied and daily application durations (n=65)		
Variable	**Category**	**n (%)**
Modality applied	Ultrasound	59 (90.8)
	TENS	59 (90.8)
	Infrared therapy	59 (90.8)
	ESWT	13 (20.0)
	Shortwave diathermy	35 (53.8)
Daily ultrasound application time	Not applied	6 (9.2)
	0-30 min	3 (4.6)
	30 min - 1 h	5 (7.7)
	1-2 h	35 (53.8)
	2-3 h	14 (21.5)
	3-4 h	1 (1.5)
	4-5 h	1 (1.5)
Daily ESWT application time	Not applied	52 (80.0)
	0-30 min	2 (3.1)
	30 min - 1 h	2 (3.1)
	1-2 h	4 (6.2)
	2-3 h	3 (4.6)
	3-4 h	2 (3.1)

Values are presented as n (%). Participants could apply more than one modality, so percentages do not sum to 100. Ultrasound, TENS, and infrared therapy were applied by exactly the same 59 participants, so these three variables are mathematically identical and are analyzed as a single variable in Table 4. Daily application time was recorded only for ultrasound and ESWT, and as ordinal bands rather than continuous minutes
ESWT: Extracorporeal shock wave therapy, TENS: Transcutaneous electrical nerve stimulation

who did not maintain an appropriate working posture reported higher QuickDASH scores (r=0.374, q=0.013), and longer sleep duration was associated with a lower PSQI global score (r=-0.350, q=0.022). Weekly working hours, rest breaks, and prolonged working in the same position showed no association with any outcome once multiple testing was accounted for (Table 3).

Professional quality of life was linked to both outcomes in the expected directions (Table 3). Compassion fatigue correlated with the QuickDASH score (r=0.512), sleep disturbances (r=0.479), use of sleep medication (r=0.467) and the QuickDASH work module (r=0.419), all at q≤0.003. Burnout correlated with the QuickDASH score (r=0.368, q=0.014) and with the use of sleep medication (r=0.318, q=0.046). Neither subscale retained a significant association with the PSQI global score after correction; none of the associations involving compassion satisfaction remained significant.

With respect to the modalities, ultrasound, TENS, and infrared therapy were applied by exactly the same participants and were therefore analyzed as a single variable. No association between the use of any modality (or its daily application time) and any outcome remained significant after false discovery rate correction (Table 4). Likewise, none of the associations between sociodemographic variables and the outcomes survived correction; the associations shown in Table 5 are exploratory and hypothesis-generating.

Factors Independently Associated with Disability and Sleep Quality

In the model for upper extremity disability, four variables were independently associated with a higher square-root–transformed QuickDASH score: upper extremity pain (B=0.476, 95% CI 0.285 to 0.667; β=0.46, p<0.001); inappropriate working posture (B=0.854, 95% CI 0.171 to 1.537; β=0.19, p=0.015); the PSQI global score (B=0.166, 95% CI 0.034 to 0.299; β=0.22, p=0.015); and years of professional experience (B=0.051, 95% CI 0.009 to 0.094; β=0.20, p=0.019). Compassion fatigue was not independently associated with disability (p=0.114). The model explained 69% of the variance [R^2=0.691, adjusted R^2=0.665, F(5, 59)=26.37, p<0.001]. Residuals met the assumptions of linearity, homoscedasticity, and normality (Shapiro-Wilk p=0.136). The Durbin-Watson statistic was 1.81, the maximum Cook's distance was 0.212, and all VIFs remained below 1.7, with tolerance values above 0.60 (Table 6).

In the model for sleep quality, upper extremity pain (B=0.521, 95% CI 0.226 to 0.816; β=0.39, p<0.001) and shorter self-reported sleep duration (B=-1.648, 95% CI -2.767 to -0.530; β=-0.30, p=0.005) were the only independent correlates of the PSQI global score. Burnout (p=0.240), compassion fatigue (p=0.581), and educational background (p=0.217) were not significant. This model explained 41% of the variance (R^2=0.412, adjusted R^2=0.363, F(5, 59)=8.28, p<0.001). Assumption checks and collinearity diagnostics are reported in Table 7.

Participants who had completed a four-year program reported higher QuickDASH scores than those who had completed a two-year program (median 21.7 versus 14.2, p=0.018) and higher PSQI global scores (median 8.0 versus 6.0, p=0.033). The same group was also considerably older and more experienced (both p<0.001); therefore, the difference is confounded by cumulative occupational exposure rather than being attributable to education itself.

Discussion

This study simultaneously examined upper extremity and neck pain, upper extremity disability, sleep quality, and professional quality of life in an occupational group that has rarely been studied in its own right. Three findings stand out. First, pain intensity was the dominant correlate of both outcome domains and remained independently associated with both disability and sleep quality after

Table 3. Correlations between pain, work-related factors, professional quality of life, sleep quality, and upper extremity disability (n=65)

Part A. Sleep quality (PSQI global score and components)

Variable		PSQI global	PSQI subjective sleep quality	PSQI sleep latency	PSQI sleep duration	PSQI habitual sleep efficiency
Upper extremity pain (NRS)	r	0.562	0.458	0.325	0.045	0.205
	p	<0.001	<0.001	0.008	0.720	0.102
	q	<0.001	0.001	0.040	0.870	0.249
Neck pain (NRS)	r	0.479	0.417	0.199	0.059	0.104
	p	<0.001	<0.001	0.111	0.639	0.410
	q	<0.001	0.003	0.250	0.847	0.644
Self-reported sleep duration	r	-0.350	-0.256	-0.235	-0.305	-0.032
	p	0.004	0.040	0.059	0.013	0.801
	q	0.022	0.125	0.159	0.056	0.914
Years of professional experience	r	0.135	0.065	0.013	0.052	-0.012
	p	0.282	0.606	0.918	0.683	0.923
	q	0.509	0.844	0.949	0.847	0.949
Weekly working hours	r	0.028	0.088	-0.020	0.038	-0.053
	p	0.822	0.485	0.876	0.766	0.676
	q	0.914	0.720	0.932	0.906	0.847
Taking rest breaks during work	r	-0.034	0.041	0.010	-0.051	-0.130
	p	0.789	0.747	0.935	0.684	0.302
	q	0.913	0.893	0.952	0.847	0.526
Prolonged working in the same position	r	-0.057	-0.005	-0.049	-0.140	-0.095
	p	0.651	0.966	0.696	0.266	0.449
	q	0.847	0.975	0.850	0.505	0.686
Maintaining appropriate body posture	r	0.125	0.137	0.068	0.092	0.136
	p	0.323	0.276	0.589	0.464	0.280
	q	0.530	0.509	0.830	0.700	0.509
ProQOL-compassion satisfaction	r	-0.251	-0.128	-0.236	-0.101	0.132
	p	0.043	0.310	0.058	0.422	0.294
	q	0.129	0.526	0.159	0.653	0.522
ProQOL-burnout	r	0.304	0.203	0.145	0.126	-0.196
	p	0.014	0.105	0.248	0.316	0.117
	q	0.056	0.249	0.479	0.526	0.258
ProQOL-compassion fatigue	r	0.303	0.239	0.013	-0.003	-0.175
	p	0.014	0.055	0.916	0.978	0.163
	q	0.056	0.158	0.949	0.978	0.346

Part B. Sleep quality (continued) and upper extremity disability

Variable		PSQI sleep disturbances	PSQI use of sleep medication	PSQI daytime dysfunction	QuickDASH	QuickDASH work module
Upper extremity pain (NRS)	r	0.420	0.434	0.433	0.769	0.764
	p	<0.001	<0.001	<0.001	<0.001	<0.001
	q	0.003	0.003	0.003	<0.001	<0.001
Neck pain (NRS)	r	0.423	0.282	0.430	0.659	0.642
	p	<0.001	0.023	<0.001	<0.001	<0.001
	q	0.003	0.085	0.003	<0.001	<0.001

Table 3. Continued

Part B. Sleep quality (continued) and upper extremity disability

Variable		PSQI sleep disturbances	PSQI use of sleep medication	PSQI daytime dysfunction	QuickDASH	QuickDASH work module
Self-reported sleep duration	r	-0.230	0.051	-0.306	0.025	-0.054
	p	0.066	0.686	0.013	0.842	0.671
	q	0.172	0.847	0.056	0.914	0.847
Years of professional experience	r	0.224	0.459	-0.085	0.325	0.281
	p	0.073	<0.001	0.500	0.008	0.023
	q	0.186	0.001	0.729	0.040	0.085
Weekly working hours	r	0.117	0.276	-0.127	0.165	0.164
	p	0.353	0.026	0.314	0.189	0.193
	q	0.563	0.093	0.526	0.381	0.381
Taking rest breaks during work	r	0.064	-0.036	-0.055	0.025	-0.080
	p	0.615	0.777	0.661	0.843	0.528
	q	0.846	0.909	0.847	0.914	0.754
Prolonged working in the same position	r	-0.031	0.238	-0.117	-0.024	0.019
	p	0.807	0.056	0.352	0.848	0.881
	q	0.914	0.158	0.563	0.914	0.932
Maintaining appropriate body posture	r	0.084	-0.026	0.053	0.374	0.251
	p	0.503	0.836	0.676	0.002	0.044
	q	0.729	0.914	0.847	0.013	0.129
ProQOL-compassion satisfaction	r	-0.163	-0.184	-0.202	-0.170	-0.200
	p	0.194	0.142	0.106	0.177	0.110
	q	0.381	0.305	0.249	0.367	0.250
ProQOL-burnout	r	0.269	0.318	0.272	0.368	0.265
	p	0.030	0.010	0.028	0.003	0.033
	q	0.100	0.046	0.097	0.014	0.107
ProQOL-compassion fatigue	r	0.479	0.467	0.214	0.512	0.419
	p	<0.001	<0.001	0.087	<0.001	<0.001
	q	<0.001	0.001	0.216	<0.001	0.003

r: Spearman rank correlation coefficient; p: unadjusted two-sided p value; q: Benjamini-Hochberg false discovery rate-adjusted p-value computed within this table. Higher PSQI and QuickDASH scores denote poorer sleep and greater disability, respectively. Dichotomous variables were coded 1= yes, 2= no (taking rest breaks, prolonged working in the same position, maintaining appropriate body posture); sleep duration was coded as an ordinal band (1= 1-4 h to 4= >9 h). A positive coefficient for maintaining appropriate body posture therefore indicates that participants who did NOT maintain an appropriate posture reported greater disability
NRS: Numeric Rating Scale, ProQOL: Professional Quality of Life Scale, PSQI: Pittsburgh Sleep Quality Index, QuickDASH: Quick disabilities of the arm, shoulder and hand questionnaire

adjustment. Second, the variables that carried independent weight for disability were cumulative rather than task-specific: years of professional experience and failure to maintain an appropriate working posture, with the latter reported by almost half of the sample. Third, psychosocial burden was clearly associated with both outcomes at the bivariate level, but it was not independently associated with either outcome once pain and workload were accounted for.

The association between pain and upper extremity disability is consistent with evidence from comparable healthcare settings. The upper extremity is among the most frequently affected regions in physiotherapists (4), and prolonged hand-arm use has been linked to musculoskeletal complaints in operating room nurses (18). Our data add a quantitative gradient within a single workforce delivering device-based treatments: the correlation between pain intensity and QuickDASH was the strongest association in the entire dataset, and pain alone accounted for the largest share of explained variance in the regression model. This supports the view that repetitive manual tasks and device-based applications

Table 4. Correlations between physical therapy modalities, application durations, sleep quality, and upper extremity disability (n=65)

Part A. Sleep quality (PSQI global score and components)

Variable		PSQI global	PSQI subjective sleep quality	PSQI sleep latency	PSQI sleep duration	PSQI habitual sleep efficiency
Use of ultrasound/TENS/infrared therapy	r	-0.068	0.119	-0.075	-0.313	-0.395
	p	0.589	0.345	0.552	0.011	0.001
	q	0.755	0.571	0.755	0.188	0.056
Use of ESWT	r	0.278	0.018	0.167	0.254	0.193
	p	0.025	0.887	0.183	0.041	0.124
	q	0.188	0.964	0.381	0.188	0.282
Use of shortwave diathermy	r	0.030	-0.019	-0.006	-0.244	0.028
	p	0.814	0.879	0.962	0.050	0.828
	q	0.940	0.964	0.982	0.201	0.940
Daily ultrasound application time	r	0.069	0.035	0.073	-0.104	-0.264
	p	0.587	0.781	0.564	0.409	0.034
	q	0.755	0.930	0.755	0.638	0.188
Daily ESWT application time	r	0.263	-0.002	0.168	0.256	0.212
	p	0.035	0.988	0.181	0.040	0.091
	q	0.188	0.988	0.381	0.188	0.266

Part B. Sleep quality (continued) and upper extremity disability

Variable		PSQI sleep disturbances	PSQI use of sleep medication	PSQI daytime dysfunction	QuickDASH	QuickDASH work module
Use of ultrasound/TENS/infrared therapy	r	0.273	0.136	-0.147	-0.069	-0.155
	p	0.028	0.282	0.244	0.582	0.218
	q	0.188	0.521	0.469	0.755	0.435
Use of ESWT	r	0.117	0.226	0.235	0.256	0.297
	p	0.354	0.071	0.060	0.039	0.016
	q	0.571	0.221	0.201	0.188	0.188
Use of shortwave diathermy	r	0.199	-0.121	0.129	-0.046	-0.052
	p	0.111	0.338	0.304	0.715	0.683
	q	0.282	0.571	0.543	0.872	0.853
Daily ultrasound application time	r	0.198	0.202	0.010	-0.008	-0.091
	p	0.114	0.107	0.934	0.947	0.471
	q	0.282	0.282	0.982	0.982	0.714
Daily ESWT application time	r	0.085	0.193	0.234	0.237	0.287
	p	0.498	0.123	0.060	0.058	0.021
	q	0.733	0.282	0.201	0.201	0.188

r: Spearman rank correlation coefficient; p: unadjusted two-sided p value; q: Benjamini-Hochberg false discovery rate-adjusted p-value computed within this table. Higher PSQI and QuickDASH scores denote poorer sleep and greater disability, respectively. Modality use was coded 0= not used, 1= used; application time was coded as an ordinal band (0= not applied to 6= 4-5 h). No association in this table remained significant after false discovery rate correction; these analyses are exploratory and hypothesis-generating

ESWT: Extracorporeal shock wave therapy, PSQI: Pittsburgh Sleep Quality Index, QuickDASH: Quick disabilities of the arm, shoulder and hand questionnaire, TENS: Transcutaneous electrical nerve stimulation

Table 5. Correlations between sociodemographic variables, sleep quality, and upper extremity disability (n=65)

Part A. Sleep quality (PSQI global score and components)

Variable		PSQI global	PSQI subjective sleep quality	PSQI sleep latency	PSQI sleep duration	PSQI habitual sleep efficiency
Age	r	0.094	0.062	0.026	0.038	-0.022
	p	0.458	0.621	0.838	0.764	0.864
	q	0.763	0.861	0.943	0.932	0.943
Sex	r	-0.120	-0.050	-0.007	-0.101	-0.051
	p	0.343	0.691	0.953	0.422	0.684
	q	0.713	0.909	0.981	0.741	0.909
Height	r	-0.114	-0.099	-0.015	0.006	0.014
	p	0.365	0.434	0.903	0.962	0.914
	q	0.713	0.748	0.952	0.981	0.952
Weight	r	-0.083	-0.041	0.039	-0.068	0.153
	p	0.513	0.748	0.757	0.593	0.224
	q	0.778	0.932	0.932	0.847	0.691
Marital status	r	-0.146	-0.150	0.001	-0.056	0.020
	p	0.245	0.231	0.994	0.658	0.877
	q	0.691	0.691	0.994	0.890	0.943
Educational attainment	r	0.267	0.274	0.141	0.022	0.179
	p	0.032	0.027	0.262	0.864	0.154
	q	0.334	0.334	0.705	0.943	0.614
Smoking status	r	-0.185	-0.175	-0.094	0.003	0.145
	p	0.139	0.163	0.458	0.980	0.249
	q	0.599	0.628	0.763	0.989	0.691
Smoking amount	r	0.157	0.109	0.083	0.024	-0.153
	p	0.212	0.386	0.511	0.851	0.223
	q	0.691	0.729	0.778	0.943	0.691
Alcohol use	r	-0.064	0.086	-0.184	-0.028	0.121
	p	0.614	0.493	0.142	0.824	0.339
	q	0.861	0.771	0.599	0.943	0.713
Alcohol amount	r	0.068	-0.087	0.185	0.028	-0.120
	p	0.591	0.490	0.139	0.824	0.339
	q	0.847	0.771	0.599	0.943	0.713

Part B. Sleep quality (continued) and upper extremity disability

Variable		PSQI sleep disturbances	PSQI use of sleep medication	PSQI daytime dysfunction	QuickDASH	QuickDASH work module
Age	r	0.114	0.412	-0.137	0.374	0.319
	p	0.367	<0.001	0.275	0.002	0.010
	q	0.713	0.065	0.705	0.094	0.238
Sex	r	-0.041	0.092	-0.264	-0.113	-0.232
	p	0.747	0.468	0.033	0.371	0.063
	q	0.932	0.767	0.334	0.713	0.422
Height	r	-0.015	-0.102	-0.172	-0.133	-0.218
	p	0.906	0.418	0.171	0.290	0.080
	q	0.952	0.741	0.633	0.713	0.473

Table 5. Continued

Part B. Sleep quality (continued) and upper extremity disability

Variable		PSQI sleep disturbances	PSQI use of sleep medication	PSQI daytime dysfunction	QuickDASH	QuickDASH work module
Weight	r	0.020	-0.061	-0.247	-0.156	-0.130
	p	0.874	0.628	0.048	0.214	0.302
	q	0.943	0.861	0.340	0.691	0.713
Marital status	r	-0.183	-0.255	-0.027	-0.307	-0.292
	p	0.144	0.041	0.832	0.013	0.018
	q	0.599	0.340	0.943	0.257	0.259
Educational attainment	r	0.210	0.365	0.080	0.297	0.222
	p	0.093	0.003	0.527	0.016	0.076
	q	0.515	0.094	0.786	0.259	0.473
Smoking status	r	-0.204	-0.252	-0.138	0.023	0.074
	p	0.103	0.043	0.272	0.859	0.559
	q	0.540	0.340	0.705	0.943	0.822
Smoking amount	r	0.191	0.250	0.103	-0.024	-0.088
	p	0.127	0.045	0.412	0.849	0.487
	q	0.599	0.340	0.741	0.943	0.771
Alcohol use	r	-0.107	0.148	-0.041	0.123	0.131
	p	0.397	0.240	0.749	0.329	0.297
	q	0.735	0.691	0.932	0.713	0.713
Alcohol amount	r	0.115	-0.148	0.046	-0.116	-0.126
	p	0.363	0.241	0.716	0.358	0.316
	q	0.713	0.691	0.930	0.713	0.713

r: Spearman rank correlation coefficient; p: unadjusted two-sided p-value; q: Benjamini-Hochberg false discovery rate-adjusted p-value computed within this table. Higher PSQI and QuickDASH scores denote poorer sleep and greater disability, respectively. Categorical variables were coded as follows: sex, 1= female, 2= male; marital status, 1= married, 2= single; educational background, 2= two-year higher education program, 4= four-year higher education programme; smoking status and alcohol use, 1= yes, 2= no; smoking and alcohol amount, 0= never, 1= occasionally, 2= frequently. No association in this table remained significant after false discovery rate correction; these analyses are exploratory and hypothesis-generating
PSQI: Pittsburgh Sleep Quality Index, QuickDASH: Quick Disabilities of the Arm, Shoulder and Hand questionnaire

compromise upper extremity functional capacity in this occupational group. The median QuickDASH score of 17.5 indicates mild to moderate disability across the sample rather than isolated severe cases, and a recent synthesis of preventive strategies for musculoskeletal disorders in physiotherapy staff has emphasized that interventions are most effective when targeted before symptoms reach the level of functional loss (19).

Our findings on exposure point away from the device and towards the organization of the working day. Neither the type of modality applied nor its daily application time distinguished participants with respect to any outcome after adjustment for multiple testing, whereas years of professional experience and working posture did. The null findings may partly reflect the relatively homogeneous exposure profile, with almost all participants delivering the same three modalities and daily application times clustering within a narrow range; however, limited statistical power cannot be excluded. This cumulative pattern contrasts with two recent meta-analyses in which years of practice were inversely related to neck and shoulder WRMSD prevalence in nurses and surgeons (1,2). Two explanations are plausible. Those syntheses report symptom prevalence, whereas our outcome was functional disability; the two do not necessarily move together. In addition, a healthy worker survivor effect may operate in larger and more mobile workforces, where affected staff transfer out of the most demanding roles; in a single small unit with limited role mobility, such selective attrition is less likely, so cumulative exposure may remain visible. Analyses in industrial populations have similarly concluded that physical and organizational workload dimensions must

Table 6. Factors associated with upper extremity disability (square-root-transformed QuickDASH score): multivariable linear regression analysis (n=65)

Variable	B	SE	95% CI for B	Standardized β	t	p	VIF/tolerance
Upper extremity pain (NRS)	0.476	0.096	0.285 to 0.667	+0.46	4.98	<0.001	1.65/0.60
Years of professional experience	0.051	0.021	0.009 to 0.094	+0.20	2.42	0.019	1.29/0.77
PSQI global score	0.166	0.066	0.034 to 0.299	+0.22	2.52	0.015	1.43/0.70
ProQOL-compassion fatigue	0.032	0.020	-0.008 to 0.073	+0.13	1.61	0.114	1.27/0.79
Maintaining appropriate body posture (1= yes, 2= no)	0.854	0.341	0.171 to 1.537	+0.19	2.50	0.015	1.16/0.86

Model statistics: R^2=0.691; adjusted R^2=0.665; F(5, 59)=26.37; p<0.001. Durbin-Watson =1.81. Shapiro-Wilk test of residuals p=0.136. Maximum Cook's distance =0.212. The model included 5 predictors, corresponding to approximately 13.0 participants per predictor.
The dependent variable was square-root transformed because the residuals of the untransformed model were not normally distributed (Shapiro-Wilk p<0.001); after transformation the assumption was met. Candidate predictors were selected on the basis of clinical relevance and bivariate associations at p<0.10. To limit overfitting, the number of predictors was capped at five, corresponding to approximately 13 participants per predictor. Neck pain and age were not entered because of collinearity with upper extremity pain (rho=0.66) and with years of professional experience (rho=0.85), respectively. Variables were entered simultaneously (enter method)
B: Unstandardized regression coefficient, CI: Confidence interval, NRS: Numeric Rating Scale, ProQOL: Professional Quality of Life Scale, PSQI: Pittsburgh Sleep Quality Index, QuickDASH: Quick disabilities of the arm, shoulder and hand questionnaire, SE: Standard error, VIF: Variance inflation factor

Table 7. Factors associated with sleep quality (PSQI global score): multivariable linear regression analysis (n=65)

Variable	B	SE	95% CI for B	Standardized β	t	p	VIF/tolerance
Upper extremity pain (NRS)	0.521	0.147	0.226 to 0.816	+0.39	3.54	<0.001	1.20/0.83
Self-reported sleep duration	-1.648	0.559	-2.767 to -0.530	-0.30	-2.95	0.005	1.07/0.94
ProQOL-burnout	0.068	0.057	-0.047 to 0.183	+0.16	1.19	0.240	1.75/0.57
ProQOL-compassion fatigue	0.024	0.043	-0.063 to 0.111	+0.07	0.55	0.581	1.79/0.56
Educational attainment	0.390	0.312	-0.236 to 1.015	+0.14	1.25	0.217	1.18/0.85

Model statistics: R^2=0.412; adjusted R^2=0.363; F(5, 59)=8.28; p<0.001. Durbin-Watson =2.39. Shapiro-Wilk test of residuals p=0.204. Maximum Cook's distance =0.162. The model included 5 predictors, corresponding to approximately 13.0 participants per predictor.
Candidate predictors were selected on the basis of clinical relevance and bivariate associations at p<0.10, with the number of predictors capped at five to limit overfitting. Neck pain was not entered because of collinearity with upper extremity pain (rho=0.66). Variables were entered simultaneously (enter method)
B: Unstandardized regression coefficient, CI: Confidence interval, NRS: Numeric Rating Scale, ProQOL: Professional Quality of Life Scale, PSQI: Pittsburgh Sleep Quality Index, QuickDASH: Quick disabilities of the arm, shoulder and hand questionnaire, SE: Standard error, VIF: Variance inflation factor

be considered jointly rather than reduced to a single task descriptor (5,7). The practical implication is that ergonomic programs for this group should target rotation, scheduled recovery between applications, workstation height, and patient positioning rather than the individual device.

The pain-sleep findings extend a well-characterized relationship to a new occupational group. A bidirectional association between sleep problems and chronic musculoskeletal pain has been established in a meta-analysis (6), and sleep quality has been proposed as a mediator between psychosocial burden and multisite musculoskeletal pain in healthcare workers (7,20). Our sample can be benchmarked directly against comparable staff: 58.5% had a PSQI global score above 5 and the median global score was 6, close to the mean of 7.0 recently reported in physiotherapists working in a physical medicine and rehabilitation clinic (8). Notably, that study also identified longer work tenure as a correlate of poorer sleep, replicating the association we observed between years of professional experience and use of sleep medication. Our data further demonstrate directional relationships within a single cohort: the PSQI global score was independently associated with upper extremity disability, while pain intensity was independently associated with the PSQI global score. This reciprocal specification should be interpreted with caution, since the PSQI global score is a predictor in one model and the outcome in the other, and the two models therefore do not provide independent confirmation of one another. A 2026 narrative review has described the neurophysiological and behavioral pathways through which the pain–sleep loop is maintained (21), and a pragmatic stepped framework for screening sleep dysfunction in musculoskeletal practice has recently been proposed (22); given that more than half of our sample screened positive, incorporating such a screen into occupational health assessment for this workforce is clinically justified. Longer self-reported sleep duration was independently associated with better sleep quality; however, because sleep duration was recorded in ordinal bands and 98% of the sample fell into just two of these bands, this association should not be overinterpreted.

The psychosocial findings require a more cautious reading than the bivariate results alone would suggest. Compassion fatigue was associated with upper extremity disability, sleep disturbances, and use of sleep medication, while burnout was associated with disability and use

of sleep medication; all associations persisted after correction for multiple testing; however, neither subscale remained independently associated with the outcomes once pain intensity and workload were entered into the models. Two interpretations are compatible with this pattern. Psychosocial burden may act largely through pain rather than alongside it, which would be consistent with reports that sleep quality partially mediates the path from burnout to multisite musculoskeletal pain in healthcare workers (20). Cross-sectional studies published in 2025 report the same clustering in nursing staff, both between sleep quality and burnout symptoms (23) and between professional quality of life measured with the ProQOL and sleep quality (24). Alternatively, the present sample may have been too small to separate the two, particularly given that the burnout subscale showed only modest internal consistency in the Turkish validation study (15). Nonetheless, the association is real at the descriptive level and warrants attention in occupational health assessment, but our data do not support the stronger claim that burnout independently determines disability or sleep quality in this group. Prospective studies with formal mediation analysis are needed to distinguish these possibilities.

Study Limitations

This study has several limitations. The cross-sectional design does not allow causal or temporal inferences; our findings are equally compatible with pain impairing sleep, poor sleep amplifying pain and disability, or a third factor driving both. The study was conducted at a single center with 65 participants and could reliably detect only medium-sized effects ($r \geq 0.34$), so weaker but clinically relevant associations may have been missed and the findings describe this workforce rather than the occupational group in general. Participation was voluntary and the total number of eligible personnel could not be retrieved retrospectively; therefore, the participation rate could not be calculated, and volunteer bias cannot be excluded. All exposures and outcomes were self-reported, without clinical examination or objective measurement of workload, which may introduce recall and social desirability biases and may inflate associations between instruments completed in the same session; in addition, the NRS captures pain intensity but not duration or chronicity. All participants worked day shifts only, so shift work could not confound the sleep findings; by the same token, the results may not extend to rehabilitation staff working rotating or night schedules. Finally, the large number of bivariate comparisons increases the risk of type I error. A false discovery rate correction was applied. Because no association in the modality or sociodemographic tables survived correction, those analyses should be regarded as exploratory and hypothesis-generating.

Despite these limitations, this study has notable strengths. To our knowledge, it is the first study to examine upper extremity pain, disability, sleep quality, and professional quality of life together in healthcare workers whose daily work consists of delivering physical therapy modalities. All participants were assessed with validated Turkish instruments under standardized conditions, and complete data were available for all analyzed participants. Both regression models were specified with an a priori cap on predictors; assumptions, collinearity diagnostics, and influence statistics were reported in full. Correction for multiple testing was applied throughout, and reporting follows the STROBE statement, which should facilitate replication elsewhere.

Conclusion

In this single-center, cross-sectional sample of healthcare workers delivering physical therapy modalities, upper extremity and neck pain intensity was the variable most consistently associated with both upper extremity disability and sleep quality, and it remained independently associated with each after adjustment. Years of professional experience and failure to maintain an appropriate working posture were independently associated with greater disability, whereas neither the modality applied nor its daily application time distinguished between participants once multiple testing was accounted for. Burnout and compassion fatigue were associated with disability and sleep disturbance at the bivariate level but were not independently associated with the outcomes alongside pain and workload. These associations are cross-sectional and do not establish causality. They suggest, however, that occupational health assessment for this workforce should focus on the organization of the working day and working posture rather than on the individual device and that a brief sleep screen may be warranted, given that more than half of the participants reported poor sleep quality. Longitudinal multicenter studies are required before any preventive strategy for this group can be recommended with confidence.

Ethics

Ethics Committee Approval: This study was approved by the University of Health Sciences Türkiye, Bakirkoy Dr. Sadi Konuk Training and Research Hospital Clinical Research Ethics Committee (approval number: 2023-15-15, date: 07.08.2023).

Informed Consent: Written informed consent was obtained from all participants.

Acknowledgments

The authors would like to thank all participants for their valuable contributions to this study.

Footnotes

Authorship Contributions

Concept: S.S., Design: S.S., Y.S.O., Data Collection or Processing: S.S., Y.S.O., Analysis or Interpretation: S.S., Y.S.O., Literature Search: S.S., Y.S.O., Writing: S.S., Y.S.O.

Conflict of Interest: The authors declared that there were no conflicts of interest.

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

Declaration Regarding the Use of AI and AI-Assisted Technologies: Artificial intelligence-assisted tools were used for language editing, manuscript formatting, table preparation, and literature checking. All outputs were reviewed and verified by the authors. The authors take full responsibility for the content of the manuscript, the accuracy of the data, and the integrity of the work.

References

1. Jacquier-Bret J, Gorce P. Work-related musculoskeletal disorder prevalence by body area among nurses in asia: systematic review and meta-analysis. Int J Environ Res Public Health. 2025;22:652.
2. Gorce P, Jacquier-Bret J. Continental assessment of work-related musculoskeletal disorders prevalence among surgeons: systematic review and meta-analysis. J Funct Morphol Kinesiol. 2025;10:221.
3. Gorce P, Jacquier-Bret J. Global prevalence of musculoskeletal disorders among physiotherapists: a systematic review and meta-analysis. BMC Musculoskelet Disord. 2023;24:265.
4. Peña-Curbelo V, Meneses-Monroy A, Mayor-Silva LI, Martín-Casas P, Álvarez-Melcón ÁC. Work-related musculoskeletal disorders in physical therapists: a cross-sectional study. J Clin Med. 2024;13:7425.
5. Govaerts R, Tassignon B, Ghillebert J, et al. Prevalence and incidence of work-related musculoskeletal disorders in secondary industries of 21st century Europe: a systematic review and meta-analysis. BMC Musculoskelet Disord. 2021;22:751.
6. Runge N, Ahmed I, Saueressig T, et al. The bidirectional relationship between sleep problems and chronic musculoskeletal pain: a systematic review with meta-analysis. Pain. 2024;165:2455-67.
7. Darvishi E, Osmani H, Aghaei A, Moloud EA. Hidden risk factors and the mediating role of sleep in work-related musculoskeletal discomforts. BMC Musculoskelet Disord. 2024;25:256.
8. Stanić M, Krasnik R, Mikov A, et al. Assessment of sleep quality among physiotherapists using the Pittsburgh Sleep Quality Index (PSQI): a pilot study. Work. 2026;83:219-26.
9. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007;370:1453-7.
10. Beaton DE, Wright JG, Katz JN; Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. J Bone Joint Surg Am. 2005;87:1038-46.
11. Koldas Dogan S, Ay S, Evcik D, Baser O. Adaptation of Turkish version of the questionnaire quick disability of the arm, shoulder, and hand (Quick DASH) in patients with carpal tunnel syndrome. Clin Rheumatol. 2011;30:185-91.
12. Buysse DJ, Reynolds CF 3rd, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28:193-213.
13. Ağargün MY, Kara H, Anlar Ö. The validity and reliability of the Pittsburgh Sleep Quality Index. Turk Psikiyatri Derg. 1996;7:107-15.
14. Stamm BH. The ProQOL manual: the Professional Quality of Life Scale: compassion satisfaction, burnout and compassion fatigue/secondary trauma scales. Pocatello: Idaho State University and Sidran Press; 2005.
15. Yeşil A, Ergün Ü, Amasyalı C, Er F, Olgun NN, Aker AT. Validity and reliability of the Turkish version of the Professional Quality of Life Scale. Archives of Neuropsychiatry 2010;47:111-7.
16. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. 1995;57:289-300.
17. Faul F, Erdfelder E, Buchner A, Lang AG. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behav Res Methods. 2009;41:1149-60.
18. Clari M, Garzaro G, Di Maso M, et al. Upper limb work-related musculoskeletal disorders in operating room nurses: a multicenter cross-sectional study. Int J Environ Res Public Health. 2019;16:2844.
19. Harithasan D, Pei Sean L, Singh DKA. Unveiling the key elements of effective preventive strategies: Insights into addressing work-related musculoskeletal disorders among physiotherapists. Work. 2025;82:130-43.
20. Vieira LMSMA, Mininel VA, Sato TO. Sleep quality as a mediator of burnout, stress and multisite musculoskeletal pain in healthcare workers: a longitudinal study. Healthcare (Basel). 2023;11:2476.
21. Newman B, Gill S, Thornton N, Hartnett J. The intersection of sleep disorders and chronic musculoskeletal pain: a narrative review. Curr Phys Med Rehabil Rep. 2026;14:24.
22. Shepherd MH, Neilson BD, Siengsukon CF. The pain of poor sleep: a clinician's guide to assessing and addressing sleep dysfunction in people with musculoskeletal pain conditions. JOSPT Open. 2026;4:12-22.
23. Baranauskas M, Kupčiūnaitė I, Lieponienė J, Stukas R. Sleep quality and professional burnout in clinical nurses: a cross-sectional study. Healthcare (Basel). 2025;13:2727.
24. Mansouri P, Rivaz M, Soltani R, Tabei P. The relationship between professional quality of life and sleep quality among nurses working in intensive care units: a cross-sectional study. BMC Nurs. 2025;24:34.

Original Article

DOI: 10.4274/haseki.galenos.2026.59455
Med Bull Haseki 2026;64(4):285-292

Urinary Iodine Concentration and Thyroid Function in Children and Adolescents: A Cross-sectional Study

Seniha Kiremitci Yilmaz[1], Ilkana Hasankhanli[2]

[1]*University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Pediatric Endocrinology, Istanbul, Türkiye*
[2]*University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Pediatrics, Istanbul, Türkiye*

Abstract

Aim: Iodine imbalance may affect thyroid function in children, but its role in suspected thyroid dysfunction in children remains unclear. This study aimed to evaluate urinary iodine status and its relationship with thyroid functional status in a pediatric endocrine referral cohort.

Methods: This retrospective cross-sectional study included 516 children and adolescents who were evaluated between April 2020 and April 2026. Urinary iodine concentration (UIC), thyroid function tests, thyroid autoantibodies, anthropometric data, pubertal status, and thyroid ultrasonography findings were reviewed. Comparisons across UIC categories were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR) procedure. Univariable and multivariable logistic regression analyses were performed to identify factors independently associated with thyroid dysfunction and iodine deficiency.

Results: Median UIC was 150 µg/L, with 34.1% of patients having UIC <100 µg/L and 8.5% having UIC ≥300 µg/L. Urinary iodine concentration categories were similarly distributed across groups with normal thyroid function, subclinical hypothyroidism, and overt hypothyroidism. Age and pubertal status differed significantly across UIC categories after FDR correction. Body mass index category was independently associated with iodine deficiency, whereas urinary iodine status was not independently associated with thyroid dysfunction.

Conclusion: In this referral-based pediatric cohort, the median UIC indicated overall iodine sufficiency, but individual iodine status varied widely. Spot UIC was not clearly related to thyroid function, thyroid autoimmunity, or ultrasonographic abnormalities. It may identify background nutritional variability but should not be used alone to explain thyroid-stimulating hormone elevation.

Keywords: Hypothyroidism, thyroid diseases, ıodine, child, adolescent

Introduction

Subclinical hypothyroidism (SCH) is defined as an elevated serum thyroid-stimulating hormone (TSH) concentration in the presence of normal circulating free thyroxine (fT4) levels (1,2). It affects approximately 3% of children and adolescents and is a common reason for referral to pediatric endocrinology clinics (3). Although the clinical course of SCH in children is often benign and may remain stable or resolve spontaneously, its etiology is heterogeneous and includes autoimmune thyroiditis, iodine imbalance, obesity, genetic factors, and transient physiological variations (1,4-6). Iodine is an essential micronutrient required for thyroid hormone synthesis, and both iodine deficiency and iodine excess may adversely affect thyroid function (7-9). Urinary iodine concentration (UIC) is widely accepted as a practical biomarker of recent iodine intake and is recommended by the World Health Organization (WHO) for assessing iodine status at the population level (7).

Alterations in iodine intake may affect thyroid physiology, including thyroid hormone production and thyroid morphology (7). Although iodine deficiency remains an important public health concern in some

Corresponding Author: Seniha Kiremitci Yilmaz, MD, University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Pediatric Endocrinology, Istanbul, Türkiye
E-mail: skyilmaz.dr@gmail.com **ORCID:** orcid.org/0000-0002-9146-4809
Received: 09.06.2026 **Accepted:** 18.08.2026 **Publication Date:** 24.09.2026

Cite this article as: Kiremitci Yilmaz S, Hasankhanli I. Urinary iodine concentration and thyroid function in children and adolescents: a cross-sectional study. Med Bull Haseki. 2026;64(4):285-292

pediatric populations, excessive iodine exposure may also be associated with thyroid dysfunction, particularly susceptible individuals (7). Previous studies of childhood SCH predominantly focused on its natural course, thyroid autoimmunity, and predictors of progression to overt hypothyroidism (1,3). However, data on iodine status and its relationship with the biochemical, clinical, and ultrasonographic characteristics of thyroid dysfunction are limited among children referred to a pediatric endocrinology clinic for suspected thyroid dysfunction. Evaluation of UIC together with thyroid function tests, thyroid autoantibodies, anthropometric characteristics, pubertal status, and thyroid ultrasonography may therefore provide a more comprehensive understanding of the potential role of iodine imbalance in these patients.

We hypothesized that iodine imbalance, reflected by low or high UIC, could be associated with biochemical and ultrasonographic features of thyroid dysfunction in children and adolescents referred for suspected thyroid dysfunction. Therefore, this study aimed to determine UICs in these patients and to investigate associations between UIC and anthropometric measurements, pubertal status, serum free triiodothyronine (fT3), fT4, and TSH levels, thyroid autoantibodies, and thyroid ultrasonographic findings. By clarifying the clinical relevance of UIC in this setting, our findings may contribute to a more appropriate interpretation of iodine status and help refine the evaluation of children presenting with TSH elevation in pediatric endocrine practice.

Materials and Methods

Compliance with Ethical Standards

This study was conducted in accordance with the principles of the Declaration of Helsinki. The study protocol was approved by the University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital Non-Interventional Clinical Research Ethics Committee (approval no: 120-2026, date: 13.05.2026).

Study Design and Participants

This single-center, retrospective, cross-sectional observational study was conducted in the pediatric endocrinology outpatient clinic. Children and adolescents referred to our clinic between April 2020 and April 2026 for evaluation of suspected thyroid dysfunction were screened for eligibility. At the index outpatient visit, all referred patients underwent a standardized evaluation, including thyroid function tests, thyroid autoantibody measurements, UIC, and thyroid ultrasonography. Patients with complete clinical, laboratory, and imaging data were included unless they met any of the predefined exclusion criteria.

Subclinical hypothyroidism was defined as an elevated serum TSH concentration with a normal serum fT4 level. Overt hypothyroidism was defined as an elevated TSH concentration accompanied by a low serum fT4 level. Patients were excluded if they had congenital hypothyroidism, were receiving levothyroxine treatment, had undergone thyroid surgery, had an acute infectious condition at the time of evaluation, had a chronic systemic disease, were using medications known to interfere with thyroid function tests, or had missing urinary iodine measurements or other essential clinical data.

Data Collection

Demographic, anthropometric, laboratory, and imaging data were obtained from medical records. The recorded variables included age, sex, pubertal status, anthropometric measurements, serum TSH, fT3, fT4, anti-thyroid peroxidase antibody (anti-TPO), anti-thyroglobulin antibody (anti-Tg), hemoglobin, mean corpuscular volume, urinary iodine and creatinine levels, and thyroid ultrasonography findings.

Anthropometric and Pubertal Assessment

Body weight was measured using a digital scale, and height was measured with a stadiometer. Body mass index (BMI) was calculated as weight in kilograms divided by the square of height in meters (kg/m^2). Anthropometric measurements were evaluated according to national age- and sex-specific reference standards and standard deviation scores were calculated using software (www.ceddcozum.com). For subgroup analyses, BMI was categorized as underweight (<5^{th} percentile), normal weight (5^{th} to <85^{th} percentile), and overweight/obesity (≥85^{th} percentile) according to age- and sex-specific BMI-for-age percentiles (10). Pubertal status was assessed according to Tanner staging (11,12).

Thyroid Examination and Ultrasonography

Thyroid enlargement on physical examination was graded according to the WHO/UNICEF/ICCID classification as grade 0 (no palpable or visible goiter), grade 1 (palpable but not visible), or grade 2 (clearly visible in the normal neck position) (13). Thyroid ultrasonography was performed using a Mindray Resona 9 ultrasound system with an L15-3WU linear probe. Thyroid gland dimensions, echogenicity, parenchymal heterogeneity, and the presence of nodules were evaluated. Thyroid gland size was assessed according to age- and sex-specific reference values, and thyroid enlargement was defined as a thyroid volume SDS >2 (14).

Laboratory Measurements

Urinary iodine concentration was measured in spot urine samples and expressed in μg/L. According to WHO criteria, iodine status was classified as severe deficiency

(<20 μg/L), moderate deficiency (20-49 μg/L), mild deficiency (50-99 μg/L), adequate intake (100-199 μg/L), intake above requirements (200-299 μg/L), or excessive intake (≥300 μg/L). For analytical comparisons, UIC was also categorized as low (<100 μg/L), normal (100-299 μg/L), and high (≥300 μg/L).

Urinary iodine concentrations were measured by inductively coupled plasma-mass spectrometry using a method accredited by the Turkish Accreditation Agency. Serum fT3, fT4, TSH, anti-TPO, and anti-Tg were analyzed on the Atellica IM1600 Immunoassay System (Siemens Atellica IM Analyzer) using competitive biotinylated rabbit polyclonal immunoassays based on direct immunochemiluminescence technology. The analytical sensitivities were 0.1 ng/dL for fT4, <0.2 pg/mL for fT3, and <0.01 mIU/L for TSH. Intra- and inter-assay coefficients of variation were ≤6% for fT4, <15% for fT3, and <5% for TSH. Thyroid autoantibody positivity was defined as the presence of anti-TPO and/or anti-Tg according to the laboratory reference ranges.

Statistical Analysis

The primary outcome was UIC among children and adolescents referred for evaluation of suspected thyroid dysfunction. Secondary analyses evaluated the associations between UIC and anthropometric characteristics, pubertal status, thyroid function tests, thyroid autoantibodies, thyroid examination findings, and thyroid ultrasonographic features. The normality of continuous variables was assessed using the Shapiro-Wilk test. Normally distributed continuous variables were presented as mean ± standard deviation, whereas non-normally distributed variables were expressed as median interquartile range, as appropriate. Continuous variables were compared using Student's t-test or the one-way ANOVA for normally distributed variables and the Mann-Whitney U test or the Kruskal-Wallis test for non-normally distributed variables. Categorical variables were compared using the Pearson chi-square test; when appropriate, the likelihood-ratio and linear-by-linear association tests were used to evaluate categorical associations and ordered trends, respectively. Thus, the principal statistical analyses comprised between-group comparisons of continuous and categorical variables and correlation analyses examining the relationships between UIC and clinical, biochemical, and ultrasonographic parameters. Correlations were assessed using Pearson or Spearman correlation coefficients, as appropriate. To account for multiple comparisons across clinical, anthropometric, biochemical, and ultrasonographic variables evaluated among UIC categories, p-values were adjusted using the Benjamini-Hochberg false discovery rate (FDR) procedure; an FDR-adjusted p-value <0.05 was considered statistically significant.

Univariable and multivariable binary logistic regression analyses were performed to identify factors associated with thyroid dysfunction, defined as SCH or overt hypothyroidism. Age was entered as a continuous variable, whereas sex, pubertal status, BMI category, urinary iodine status, thyroid autoantibody positivity, and thyroiditis were entered as categorical variables. Urinary iodine status was categorized as <100, 100-299, and ≥300 μg/L. Results were expressed as odds ratios (ORs) with 95% confidence intervals (CIs). Multivariable binary logistic regression analysis was also performed to identify factors independently associated with iodine deficiency. Age was entered as a continuous variable, whereas BMI category, pubertal status (prepubertal vs. pubertal), sex, thyroid autoantibody positivity, and thyroiditis were entered as categorical variables. Analyses were performed on an available-case basis for each variable; therefore, denominators varied across comparisons. A two-sided p-value <0.05 was considered statistically significant. Statistical analyses were performed using IBM SPSS Statistics, version 22.0 (IBM Corp., Armonk, NY, USA).

Results

Patient Characteristics

A total of 739 children and adolescents were screened for eligibility. Of these, 516 were included in the final analysis (Figure 1). The median age was 10 (6-13) years; 300 (58%) were female and 216 (42%) were male; 267 (51.7%) were pubertal (Table 1).

Thyroid Function and Autoantibody Status

At initial evaluation, 282 (54.6%) patients had normal thyroid function, 191 (37.0%) had SCH, and 43 (8.3%) had overt hypothyroidism. Thyroid autoantibody positivity differed significantly across thyroid function groups and was lowest in patients with SCH (16%) and highest in those with overt hypothyroidism (30%) (p=0.042). Autoantibody-positive patients were older (p<0.001) and had lower fT4 levels (p=0.011), whereas UIC (p=0.706) and TSH levels (p=0.166) were comparable between autoantibody-positive and autoantibody-negative patients. Thyroid autoantibody positivity was also similar across UIC categories (p=0.742) (Table 1).

Thyroid Examination and Ultrasonographic Findings

Assessment of goiter was available for 465 patients; goiter was present in 73/465 (15.7%). Thyroid ultrasonography was available for 437 patients: thyroid enlargement was detected in 40 (9.1%) patients, thyroiditis-related findings were detected in 165 patients (37.7%), and thyroid nodules were detected in 51 (11.6%) patients. Ultrasonographic findings did not differ significantly across UIC categories (Table 1).

Urinary Iodine Status

The median UIC was 150 μg/L. Overall, 176 (34.1%) patients had iodine deficiency (UIC <100 μg/L), 296 (57.2%) had UIC values of 100-299 μg/L, and 44 (8.5%) had excessive iodine status (UIC ≥300 μg/L). Urinary iodine concentration categories were similarly distributed across the normal thyroid function, SCH, and overt hypothyroidism groups (p=0.801) (Table 2).

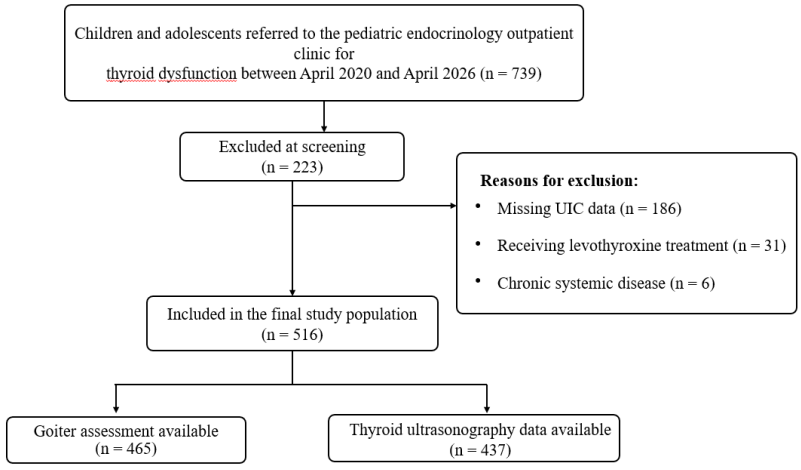

Figure 1. Study flow diagram
UIC: Urinary iodine concentration

Table 1. Clinical and biochemical characteristics according to grouped UIC categories

Variable	UIC<100 μg/L (n=176)	UIC 100-299 μg/L (n=296)	UIC ≥300 μg/L (n=44)	Raw p-value	FDR-adjusted p-value
Age, years	10.1 (6-13)	10.4 (7-14)	6.9 (1-12)	<0.001	<0.010
Sex, female/male, n	103/73	165/129	30/14	0.314	0.449
Pubertal status, prepubertal/pubertal, n	92/84	128/167	29/15	0.009	0.045
BMI SDS	0.41±1.41	0.22±1.33	-0.15±1.72	0.043	0.108
Height SDS	-0.08±1.15	-0.13±1.11	-0.18±1.16	0.966	0.966
TSH, mIU/L	5.13±4.53	4.85±4.17	5.07±2.91	0.216	0.432
fT4, ng/dL	12.0±1.8	12.6±4.2	13.0±2.4	0.034	0.108
Thyroid autoantibody positivity, n	37/153	60/267	7/38	0.742	0.824
Thyroid enlargement positivity, n	18/154	20/251	2/32	0.430	0.538
Thyroid nodule positivity, n	19/154	31/251	1/32	0.294	0.449

Data are presented as median (IQR), mean ± SD, or n, as appropriate. Denominators vary across variables because of missing data. Continuous variables were compared using one-way ANOVA or the Kruskal-Wallis test, as appropriate according to data distribution, and categorical variables were compared using Pearson's chi-square test. p-values for comparisons across UIC categories were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate procedure. An FDR-adjusted p-value <0.05 was considered statistically significant
BMI: Body-mass index, FDR: False discovery rate, fT4: Free thyroxine, IQR: Interquartile range, SDS: Standard deviation score, TSH: Thyroid stimulating hormone, UIC: Urinary iodine concentration, ANOVA: Analysis of variance

Table 2. Urinary iodine concentration categories and median UIC across thyroid function group

Thyroid function group	UIC<100 μg/L n (%)	UIC 100-299 μg/L n (%)	UIC ≥300 μg/L n (%)	p[1]-value	Median UIC, μg/L (IQR)	p[2]-value
Normal thyroid function (n=282)	95 (33.8)	165 (58.7)	21 (7.5)	0.801	145 (81-182)	0.606
SCH (n=191)	64 (33.5)	108 (56.5)	19 (9.9)		156 (83-185)	
Overt hypothyroidism (n=43)	17 (39.5)	22 (51.2)	4 (9.3)		158 (70-196)	

Data are presented as n (%) or median (IQR), as appropriate. [1]The distribution of UIC categories across thyroid function groups was compared using Pearson's chi-square test. [2]Median UIC values across thyroid function groups were compared using the Kruskal-Wallis test. A two-sided p-value <0.05 was considered statistically significant
IQR: Interquartile range, SCH: Subclinical hypothyroidism, UIC: Urinary iodine concentration

Urinary iodine levels were comparable between females and males. Patients with UIC ≥300 μg/L were younger and were frequently prepubertal. Age and pubertal status remained significantly different across UIC categories after FDR correction (adjusted p<0.010 and p=0.045, respectively) (Table 1). Body mass index category was not significantly associated with UIC category, according to Pearson's chi-squared test (p=0.070). Urinary iodine concentration values decreased across BMI categories, from 186 (100-206) μg/L in underweight children to 156 (85-183) μg/L in normal-weight children and to 132 (68-170) μg/L in children with overweight/obesity; however, this difference did not remain statistically significant after FDR correction (raw p=0.034, adjusted p=0.108).

Multivariable binary logistic regression analysis showed that BMI category was independently associated with iodine deficiency (overall p=0.010). Compared with children who were overweight or obese, underweight children (OR=0.338, 95% CI: 0.135-0.847, p=0.021) and normal-weight children (OR=0.484, 95% CI: 0.288-0.813, p=0.006) had lower odds of iodine deficiency. Age, pubertal status, sex, thyroid autoantibody positivity, and thyroiditis were not independently associated with iodine deficiency (all p>0.05).

For analytical comparison, UIC was also categorized as low (<100 μg/L), normal (100-299 μg/L), and high (≥300 μg/L). Mean fT4 levels differed across UIC categories in the unadjusted analysis (p=0.034); however, this difference was no longer statistically significant after FDR correction (adjusted p=0.108) (Table 1). No significant differences were observed in sex, height SDS, fT3, TSH, anti-TPO, anti-Tg, and thyroid volume across UIC groups. Similarly, thyroid enlargement, ultrasonographic findings of thyroiditis, and presence of thyroid nodules were comparable among iodine categories (Table 1). Urinary iodine concentration was not significantly correlated with thyroid function parameters, thyroid autoantibodies, thyroid ultrasonographic findings, height SDS, or BMI SDS.

A univariable binary logistic regression analysis showed that age and pubertal status were significantly associated with thyroid dysfunction. Each one-year increase in age was associated with a 9.7% decrease in the odds of thyroid dysfunction (OR=0.903, 95% CI: 0.869-0.938, p<0.001), while pubertal children had significantly lower odds than prepubertal children (OR=0.451, 95% CI: 0.317-0.643, p<0.001). Sex, thyroid autoantibody positivity, thyroiditis, UIC, BMI category, and goiter were not significantly associated with thyroid dysfunction (all p>0.05). However, in the multivariable binary logistic regression model, age, pubertal status, sex, urinary iodine status, thyroid autoantibody positivity, thyroiditis, and BMI category were not independently associated with thyroid dysfunction. After adjustment for other covariates, urinary iodine status was not significantly associated with thyroid dysfunction (overall p=0.410).

Discussion

Although the median UIC in our cohort was 150 μg/L, indicating iodine sufficiency at the population level according to WHO criteria, the distribution of iodine status was clearly heterogeneous (15). More than one-third of the patients had UIC <100 μg/L, whereas 8.5% had UIC ≥300 μg/L. This coexistence of iodine deficiency and excess, despite an adequate population-level median, represents one of the main findings of our study and emphasizes that a sufficient median UIC does not imply uniform iodine sufficiency at the individual level. Compared with Turkish data, our median UIC was within the range previously reported in pediatric populations (107-175 μg/L) (16-18). A similar heterogeneous pattern was recently reported by Baladastian et al. (19), who observed both iodine deficiency (19.6%) and excessive iodine status (14.0%) despite overall adequate iodine nutrition (median UIC of 190 μg/L). These findings suggest that contemporary pediatric populations may exhibit substantial variability in iodine exposure even in settings considered iodine-sufficient.

The iodine profile observed in our cohort should also be interpreted in the context of the clinically selected population. International studies have reported considerable variation in pediatric UIC across iodine-sufficient and high-exposure settings (20). However, direct comparison of absolute UIC values across populations may be of limited value because dietary iodine sources, iodized salt use, supplement exposure, age distribution, and sampling characteristics differ substantially between studies. More importantly, our cohort consisted of children referred for suspected thyroid dysfunction, rather than being a population-based sample. Thus, the heterogeneity observed in our study may reflect both background iodine exposure and the characteristics of children referred to pediatric endocrinology practice.

A clinically important observation was that the heterogeneous iodine distribution did not translate into a corresponding difference in thyroid function status. Children with normal thyroid function, SCH, and overt hypothyroidism showed broadly similar distributions across iodine categories, suggesting that neither iodine deficiency nor iodine excess alone adequately explains mild thyroid dysfunction in this referral population. This finding differs from some pediatric studies in which either excessive or insufficient iodine exposure was associated with thyroid dysfunction. Yang et al. (21) reported associations between prolonged high iodine exposure and increased thyroid volume and hyperthyrotropinemia, while Korean and Nepalese studies also suggested relationships

between iodine excess and thyroid dysfunction (22,23). Iodine deficiency also remains a recognized cause of acquired hypothyroidism in children, particularly in iodine-deficient settings (6). The differing findings across studies suggest that the thyroid response to iodine may depend not only on the amount of exposure but also on the duration of exposure, background iodine status, individual susceptibility, and population characteristics. A single spot UIC measurement in our study may, therefore, capture recent exposure without fully reflecting the longer-term iodine environment relevant to thyroid dysfunction.

In our cohort, children with excessive UIC were younger, and higher-UIC groups had a larger proportion of prepubertal children. Importantly, differences in age and pubertal status across UIC categories remained significant after correction for multiple tests. Age-related differences in diet, parent-directed feeding, nutritional supplement use, body size, fluid turnover, and urinary iodine handling may contribute to variation in spot UIC. Previous studies have reported different age-related patterns, including greater iodine deficiency in younger children and increasing iodine excess with age (24,25). This inconsistency supports the interpretation that the relationship between age and UIC is population-specific rather than universal. In clinical practice, age and pubertal stage may therefore be important contextual factors when interpreting spot UIC values in children.

The findings related to BMI require a similar interpretation. Although BMI-related differences across UIC groups were not statistically significant, the BMI category remained independently associated with iodine deficiency. Children with overweight or obesity had higher odds of iodine deficiency than normal-weight or underweight children. Children with overweight/obesity may differ from their normal-weight or underweight peers in dietary composition, consumption of iodine-containing foods, supplement use, and hydration patterns, all of which may influence spot UIC. Some pediatric studies have similarly reported lower UIC among children with overweight/obesity (26), whereas other studies, including recent data, have not identified BMI as an independent predictor of low UIC (19,24). The inconsistency across populations, together with the cross-sectional nature of our data, argues against a direct causal interpretation and suggests that BMI may instead act as a marker of differences in diet, lifestyle, or iodine exposure.

Our analyses further strengthen the conclusion that spot UIC has limited explanatory value for thyroid dysfunction in this cohort. This is consistent with studies showing that spot UIC does not necessarily result in abnormal thyroid function tests in children (27,28).

The lack of a link between iodine status and thyroid autoimmunity suggests that recent iodine exposure was not a significant factor in thyroid abnormalities in this cohort. Thyroid autoantibody levels and thyroiditis-related ultrasonographic findings were not clearly related to UIC. Previous pediatric data have similarly failed to demonstrate a consistent association between iodine status and thyroid autoimmunity (29), although iodine excess has been associated with autoimmune thyroid disease in other settings (30,31). One possible explanation is that autoimmune thyroid disease reflects longer-term interactions between genetic susceptibility, immune mechanisms, and environmental exposures, whereas a single spot UIC measurement reflects only recent iodine intake. Therefore, the absence of an association in our study does not exclude a potential effect of prolonged iodine exposure on thyroid autoimmunity.

The ultrasonographic findings should be interpreted within the same temporal framework. Structural thyroid changes are likely to reflect cumulative influences of growth, developmental stage, autoimmunity, and longer-term environmental exposures rather than short-term fluctuations in iodine intake. Consequently, the absence of a relationship between spot UIC and thyroid enlargement, thyroiditis-related changes, or nodules in our cohort is biologically plausible. In healthy Turkish school-age children from an iodine-sufficient region, thyroid volume was reported to correlate more strongly with age and anthropometric variables than with UIC itself (32). More recently, Pang et al. (9) demonstrated the relevance of anthropometric characteristics to thyroid morphology in a large iodine-sufficient pediatric population. Together, these observations suggest that a single UIC measurement may have limited ability to explain structural thyroid findings in children.

From a clinical perspective, our findings argue against using a single spot UIC value as a direct explanation for mild TSH elevation in children referred to pediatric endocrinology. Rather, UIC appears to provide contextual information about recent iodine exposure. The coexistence of deficiency and excess in the same referral population may justify considering iodine exposure when evaluating selected patients, but such interpretation should be integrated with thyroid function tests, thyroid autoantibodies, clinical findings, dietary and supplement history, and, when indicated, ultrasonographic findings. The lack of an independent association between iodine status and thyroid dysfunction further supports this comprehensive, rather than UIC-centered, approach.

Study Limitations

The retrospective cross-sectional design of the study precluded causal inference, and the single-center setting may limit the generalizability of the findings. In addition, spot UIC reflects recent iodine intake and shows

substantial intra-individual variability, limiting individual-level interpretation. Dietary iodine intake, supplement use, iodized salt exposure, and other potential sources of iodine were not systematically assessed. These unmeasured exposures are particularly relevant to the observed associations with age, pubertal status, and BMI, and they limit our ability to determine the mechanisms underlying these relationships. Moreover, the relatively small number of children with excessive iodine status may have limited the ability to detect associations within this subgroup.

Despite these limitations, the study has several strengths. The relatively large pediatric referral cohort allowed simultaneous assessment of iodine status, thyroid function, thyroid autoimmunity, anthropometric characteristics, pubertal stage, and thyroid ultrasonographic findings within the same clinical population. In addition, correction for multiple testing reduced the likelihood of interpreting chance findings as meaningful associations, while multivariable regression analyses allowed the principal relationships to be evaluated after accounting for relevant covariates. Together, these methodological features provide a more conservative assessment of the clinical relevance of spot UIC in children referred for suspected thyroid dysfunction.

Conclusion

The median spot UIC was consistent with overall iodine sufficiency in this referral-based pediatric cohort, despite substantial interindividual variability in iodine status. Importantly, this heterogeneity in iodine exposure was not associated with thyroid dysfunction, thyroid autoimmunity, or ultrasonographic abnormalities after adjustment for relevant covariates. Age and developmental stage appear to influence the distribution of spot UIC, while the association between BMI and iodine deficiency may reflect more complex dietary or behavioral factors. Therefore, a single spot UIC measurement should be interpreted primarily as a marker of recent iodine exposure and should not be used in isolation to explain TSH elevation in individual children. Its clinical value is likely greatest when integrated into a broader assessment rather than considered as an isolated diagnostic indicator.

Ethics

Ethics Committee Approval: The study protocol was approved by the University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital Non-Interventional Clinical Research Ethics Committee (approval no: 120-2026, date: 13.05.2026).

Informed Consent: Because of the retrospective design of the study, written informed consent was not obtained from the patients or their parents/legal guardians.

Authorship Contributions

Surgical and Medical Practices: S.K.Y., I.H., Concept: S.K.Y., Design: S.K.Y., Data Collection or Processing: I.H., Analysis or Interpretation: S.K.Y., Literature Search: S.K.Y., Writing: S.K.Y., I.H.

Conflict of Interest: No conflicts of interest were declared by the authors.

Financial Disclosure: This study received no financial support.

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

References

1. Lazarus J, Brown RS, Daumerie C, Hubalewska-Dydejczyk A, Negro R, Vaidya B. 2014 European Thyroid Association guidelines for the management of subclinical hypothyroidism in pregnancy and in children. Eur Thyroid J. 2014;3:76-94.
2. Cooper F, Guadalupe Rios G, Rivera-Sepulveda A, DiLeonardo O, Carakushansky M, Gurnurkar S. Subclinical but significant? Updated review of pediatric hypothyroidism. J Pediatr Endocrinol Metab. 2025;38:1247-57.
3. Lazar L, Frumkin RB, Battat E, Lebenthal Y, Phillip M, Meyerovitch J. Natural history of thyroid function tests over 5 years in a large pediatric cohort. J Clin Endocrinol Metab. 2009;94:1678-82.
4. Demirbilek H, Kandemir N, Gonc EN, Ozon A, Alikasifoglu A, Yordam N. Hashimoto's thyroiditis in children and adolescents: a retrospective study on clinical, epidemiological and laboratory properties of the disease. J Pediatr Endocrinol Metab. 2007;20:1199-205.
5. Metwalley KA, Farghaly HS. Subclinical hypothyroidism in children: updates for pediatricians. Ann Pediatr Endocrinol Metab. 2021;26:80-5.
6. Salerno M, Capalbo D, Valenzise M, et al. Unveiling the hidden: acquired pediatric hypothyroidism. Front Endocrinol (Lausanne). 2026;16:1744280.
7. World Health Organization. Urinary iodine concentrations for determining iodine status in populations. Geneva; 2013 30 August 2013. Contract No.: WHO/NMH/NHD/EPG/13.1.
8. Delić T, Karanović Štambuk S. Iodine in health and disease: a comprehensive review. Nutrients. 2026;18:1262.
9. Pang P, Xie J, Yang M, Zhou S, Zhang Y. A cross-sectional study of iodine nutritional status among school-age children in Chongqing, China. Nutrients. 2025;17:817.
10. BMI. Child and teen BMI categories. [updated 28 June 2024] Available from:s URL:https://www.cdc.gov/bmi/child-teen-calculator/bmi-categories.html?utm_source
11. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. Arch Dis Child. 1969;44:291-303.
12. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. Arch Dis Child. 1970;45:13-23.
13. WHO/UNICEF/ICCIDD. Indicators for assessing Iodine deficiency disorders and their control programmes. Geneva: World Health Organization; 1993.

14. Aydıner Ö, Karakoç Aydıner E, Akpınar İ, Turan S, Bereket A. Normative data of thyroid volume-ultrasonographic evaluation of 422 subjects aged 0-55 years. J Clin Res Pediatr Endocrinol. 2015;7:98-101.
15. Gorstein J, Sullivan KM, Parvanta I, Begin F. Indicators and methods for cross-sectional surveys of vitamin and mineral status of populations. The Micronutrient Initiative (Ottawa) and the Centers for Disease Control and Prevention (Atlanta). 2007.
16. Erdoğan MF, Ağbaht K, Altunsu T, et al. Current iodine status in Turkey. J Endocrinol Invest. 2009;32:617-22.
17. Cevik K, Aylanc H. Determination of urinary iodine excretion and iodine deficiency in school-age children aged 7-11 after mandatory iodization in Canakkale, Turkiye. North Clin Istanb. 2025;12:277-84.
18. Çelmeli G, Çürek Y, Özen Küçükçetin İ, et al. The results of 16 years of iodization: assessment of iodine deficiency among school-age children in Antalya, Turkey. J Clin Res Pediatr Endocrinol. 2020;12:256-60.
19. Baladastian S, Bahreini F, Nouri A, Morvaridi M, Tavakoli H. Iodine deficiency and excess in children and related factors in Alborz Province: a cross-sectional study. BMC Nutr. 2026;12:47.
20. Iodine Global Network. Global scorecard of iodine nutrition in 2025. Available from:s URL: https://ign.org/scorecard/
21. Yang R, Lv D, Wang X, et al. Assessment of the association of high-iodine exposure in drinking water, iodine intake, and thyroid function in 3-14-year-old children. Sci Rep. 2025;15:36625.
22. Kang MJ, Hwang IT, Chung HR. Excessive iodine intake and subclinical hypothyroidism in children and adolescents aged 6-19 years: results of the sixth Korean National Health and Nutrition Examination Survey, 2013-2015. Thyroid. 2018;28:773-9.
23. Shrestha U, Gautam N, Agrawal KK, Jha AC, Jayan A. Iodine status among subclinical and overt hypothyroid patients by urinary iodine assay: a case-control study. Indian J Endocrinol Metab. 2017;21:719-23.
24. A Cesar J, S Santos I, E Black R, Chrestani MAD, Duarte FA, Nilson EAF. Iodine status of brazilian school-age children: a national cross-sectional survey. Nutrients. 2020;12:1077.
25. Shan X, Liu C, Luo X, et al. Iodine nutritional status and related factors among chinese school-age children in three different areas: a cross-sectional study. Nutrients. 2021;13:1404.
26. De Angelis S, Bagnasco M, Moleti M, et al. Obesity and monitoring iodine nutritional status in schoolchildren: is body mass ındex a factor to consider? Thyroid. 2021;31:829-40.
27. Wallborn T, Vogel M, Kneuer A, et al. Spot urine iodine levels below the WHO recommendation are not related to impaired thyroid function in healthy children and adolescents. Eur J Nutr. 2021;60:493-502.
28. Aarsland TE, Aakre I, Stea TH, et al. Association of mild-to-moderate iodine deficiency with thyroid function-a systematic review and meta-analysis. Adv Nutr. 2025;16:100471.
29. Li X, Zhang J, Ding H, et al. Iodine nutritional status and thyroid autoimmunity in chinese children and adolescents aged 6-17 years. Nutrients. 2024;16:3720.
30. Sang Z, Chen W, Shen J, et al. Long-term exposure to excessive iodine from water is associated with thyroid dysfunction in children. J Nutr. 2013;143:2038-43.
31. Zois C, Stavrou I, Svarna E, Seferiadis K, Tsatsoulis A. Natural course of autoimmune thyroiditis after elimination of iodine deficiency in northwestern Greece. Thyroid. 2006;16:289-93.
32. Deveci Sevim R, Gök M, Öztürk S, et al. Thyroid volume in Turkish school-age children living in an iodine-sufficient region. J Pediatr Endocrinol Metab. 2024;37:228-35.

Original Article

DOI: 10.4274/haseki.galenos.2026.96977
Med Bull Haseki 2026;64(4):293-303

Clinical Spectrum and Surgical Relevance of Abdominal CT Findings in Pediatric Surgery Patients in a Tertiary Center: A Retrospective Observational Study

Sevgi Buyukbese Sarsu[1], Rustem Berhan Pirimoglu[2], Merve Ucar[3]

[1]*University of Health Sciences Türkiye, Gaziantep City Hospital, Clinic of Pediatric Surgery, Gaziantep, Türkiye*
[2]*University of Health Sciences Türkiye, Gaziantep City Hospital, Clinic of Pediatric Radiology, Gaziantep, Türkiye*
[3]*University of Health Sciences Türkiye, Gaziantep City Hospital, Clinic of Radiology, Gaziantep, Türkiye*

Abstract

Aim: Abdominal computed tomography (CT) is widely used in pediatric surgical practice despite concerns regarding radiation exposure and variability in diagnostic performance. This study aimed to characterize abdominal CT findings and their relationships with clinical indications, demographic characteristics, and surgical intervention in a tertiary pediatric surgical center.

Methods: This single-center retrospective observational study included all consecutive pediatric patients who underwent abdominal CT at a tertiary pediatric surgery center between July 1 and November 1, 2025. Demographic characteristics, clinical indications, CT findings, and surgical outcomes were analyzed. CT findings were categorized according to predefined diagnostic groups, and their association with surgical intervention was evaluated using appropriate statistical methods. Diagnostic performance measures, including sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and odds ratios (ORs) with 95% confidence intervals, were calculated where applicable.

Results: Among 401 pediatric patients, 94.0% demonstrated at least one pathological finding, while 6.0% had normal results. The mean age was 11.5±4.5 years, and 54.1% were male. The most common indications were suspected appendicitis, trauma, and pelvic/adnexal pathologies, with CT findings predominantly consisting of non-specific findings and appendiceal, trauma-related, and pelvic/adnexal pathologies. Computed tomography positivity was significantly associated with surgical intervention (30.0% vs. 4.2%, OR=9.84, p=0.007). The PPV and NPV of CT positivity for surgical intervention were 30.0% and 95.8%, respectively. The overall surgical intervention rate was 28.4%. Trauma cases were mostly blunt, with limited need for surgery (5.3%). Incidental findings were observed in 28.9% of patients. Mass lesions were uncommon and were most frequently located in the ovaries and liver.

Conclusion: Abdominal CT is commonly used for suspected appendicitis, trauma, and presentations related to pelvic or adnexal pathology. Despite a high CT positivity rate, the relatively low rate of surgical intervention indicates that CT functions as a complementary rather than a definitive tool in surgical decision-making. The high NPV suggests that negative CT findings were generally associated with non-operative management in clinically equivocal cases.

Keywords: Tomography, X-ray computed, ultrasonography, abdominal pain, child, tertiary care centers

Corresponding Author: Sevgi Buyukbese Sarsu, MD, Independent Researcher, Gaziantep, Türkiye
E-mail: sarsubuyukbese@gmail.com **ORCID:** orcid.org/0009-0004-5561-0014
Received: 14.01.2026 **Accepted:** 27.07.2026 **Publication Date:** 24.09.2026

Cite this article as: Buyukbese Sarsu S, Pirimoglu RB, Ucar M. Clinical spectrum and surgical relevance of abdominal CT findings in pediatric surgery patients in a tertiary center: a retrospective observational study. Med Bull Haseki. 2026;64(4):293-303

Introduction

Abdominal complaints in pediatric patients frequently present with diagnostic uncertainty and pose major obstacles to clinical decision-making due to their broad and age-dependent differential diagnosis (1). Delayed diagnosis of acute appendicitis in children is associated with increased morbidity and worse clinical outcomes, including higher rates of perforation and prolonged hospital stay, highlighting the clinical importance of timely and accurate diagnostic evaluation (2). According to current evidence-based guidelines, including the American College of Radiology (ACR) recommendations, ultrasonography (USG) is the first-line imaging modality for evaluating suspected abdominal pathology in pediatric patients (3).

Computed tomography (CT) should be reserved for selected cases in which USG findings are inconclusive, as supported by recent literature on pediatric imaging optimization (4). However, the diagnostic performance of USG may be limited due to operator dependency and interference from intestinal gas (5,6).

In such cases, CT allows rapid, comprehensive assessment of intra-abdominal pathology and aids in the detection of complications, particularly in clinically complex scenarios (7). CT has been shown to be an effective diagnostic modality in the evaluation of acute abdominal pain within structured diagnostic pathways (8). Additionally, CT plays a key role in evaluating and grading intra-abdominal injuries, especially in hemodynamically stable pediatric trauma patients (7). Given the potential risks associated with ionizing radiation in pediatric patients, CT should be used judiciously and performed in accordance with the ALARA principle, as its application has been shown to reduce unnecessary radiation exposure and support the use of alternative imaging modalities when appropriate (9). Furthermore, CT utilization varies significantly across institutions, reflecting differences in patient characteristics, clinical indications, and adherence to imaging guidelines, underscoring the need for center-specific evaluation and optimization of imaging practices (10). However, real-world patterns of CT utilization in tertiary pediatric surgical centers and their relationship with surgical decision-making remain insufficiently characterized.

We hypothesized that CT utilization in a tertiary pediatric surgical center reflects the complexity of referred cases and that CT findings are associated with surgical decision-making processes. Accordingly, the primary objective of this study was to characterize the spectrum of abdominal CT findings in a tertiary pediatric surgery center. The secondary objective was to evaluate the association with surgical intervention and demographic and clinical variables. This approach is expected to contribute to a more rational and evidence-based use of CT in pediatric surgical practice.

Materials and Methods

Compliance with Ethical Standards

This study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki. Ethical approval was obtained from the University of Health Sciences Türkiye, Gaziantep City Hospital Non-Interventional Clinical Research Ethics Committee (approval no: 348/2025, date: 17.12.2025). Given the retrospective design of the study, the requirement for informed consent was waived by the committee. All data were anonymized, and personal identifiers were removed.

Study Design

The study was designed and reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology guidelines for observational studies.

This single-center retrospective observational study was conducted at a tertiary pediatric surgery center (University of Health Sciences Türkiye, Gaziantep City Hospital). All consecutive pediatric patients aged 0-17 years who underwent abdominal CT between July 1 and November 1, 2025, were initially assessed for eligibility. Patients were included regardless of indication, clinical presentation, or CT findings, ensuring consecutive inclusion and reducing sampling bias.

A stepwise imaging approach generally consistent with contemporary pediatric imaging recommendations, including the 2024 AAP guidance, was applied in routine clinical practice. Computed tomography was performed selectively when USG was inconclusive (defined as inadequate visualization of the target organ or discordance between sonographic findings and the clinical presentation), when it was technically limited, or when the examination could not be completed due to body habitus or patient non-cooperation. General CT indications included clinical deterioration or persistent symptoms despite initial evaluation, peritoneal signs (guarding, rebound tenderness), and elevated inflammatory markers (CRP elevation and leukocytosis), all suggesting significant intra-abdominal pathology.

For suspected appendicitis, CT was considered when USG was inconclusive, particularly when clinical findings raised concern for complicated appendicitis or remained discordant with sonographic findings. In trauma patients, CT indications included high-energy mechanism, hemodynamic instability, positive FAST examination, abdominal wall ecchymosis (seat belt sign), elevated liver enzymes (aspartat aminotransferaz >200 IU/L or alanin aminotransferaz >125 IU/L), hematuria (≥50 red blood cell/hpf), hematocrit below 30%, and thoracic wall trauma

with abnormal chest radiography. For a suspected intra-abdominal mass or tumor, CT was indicated when cross-sectional imaging was required for lesion characterization, staging, or surgical planning. In female patients presenting with acute pelvic pain, CT was performed when non-gynecological causes were suspected, in accordance with the recommendations of the European Society of Emergency Radiology.

For the remaining indications, CT was applied as follows: bowel obstruction/ileus, when ischemia, strangulation, or closed-loop obstruction was suspected; mesenteric lymphadenitis, when an alternative diagnosis was suspected or symptoms failed to resolve; gastrointestinal bleeding, when endoscopy was unavailable, contraindicated, or inconclusive; pancreatitis, when clinical severity markers or complications were suspected; urolithiasis, as a second-line modality, when obstruction or infection was suspected; hepatobiliary pathology and acute cholangitis, when ultrasound was equivocal and clinical suspicion persisted; gastrointestinal inflammatory conditions, including enteritis and colitis, when complications such as perforation or abscess were suspected; constipation, when structural or obstructive pathology or complications such as stercoral colitis, ischemia, or perforation were clinically suspected; and non-specific abdominal pain, when ultrasound was inconclusive. For heterogeneous "other" indications, CT was applied selectively when the initial clinical and sonographic evaluations were insufficient to establish a diagnosis or to exclude significant pathology. Incidental findings were documented separately and did not constitute independent CT indications.

Patient Selection

Of the 414 patients initially assessed for eligibility, 13 were excluded: 7 due to missing clinical or radiological data and 6 due to the absence of a documented pediatric surgical consultation. Patients without documented pediatric surgical consultations were excluded because surgical consultation records constituted an essential component of the clinical dataset used for outcome classification and management assessment. The final study cohort consisted of 401 patients (Figure 1).

Classification of Abdominal CT Findings

Abdominal CT findings were classified into two primary categories: normal (CT^-), defined as the complete absence of identifiable pathological findings, and pathological (CT^+), defined as the presence of any abdominal abnormality detectable on CT, including inflammatory, traumatic, obstructive, neoplastic, or urogenital conditions.

Clinical Subgrouping of CT Findings

Computed tomography-positive findings were further stratified into two clinically meaningful subgroups based on their typical management pathway in contemporary pediatric practice. The first subgroup—surgically relevant findings—encompassed conditions associated with a high likelihood of operative intervention, including perforated appendicitis, bowel obstruction (e.g., intussusception, volvulus), ovarian torsion, significant traumatic solid organ injuries [American Association for the Surgery of Trauma (AAST) Grade III or higher, or any grade with hemodynamic instability], and intra-abdominal neoplastic lesions.

The second subgroup—conservatively managed findings—consisted of pathologies typically amenable to nonoperative management according to current pediatric clinical guidelines, including mesenteric lymphadenitis, low-grade solid organ injuries (AAST Grade I-II) without hemodynamic compromise, urinary tract infections without obstructive uropathy, non-specific inflammatory changes, and pelvic inflammatory conditions.

This two-tiered classification was predefined prior to data extraction and was applied using final radiology reports and clinical records.

The classification framework was developed in accordance with contemporary pediatric surgical and emergency medicine management principles and finalized before data extraction.

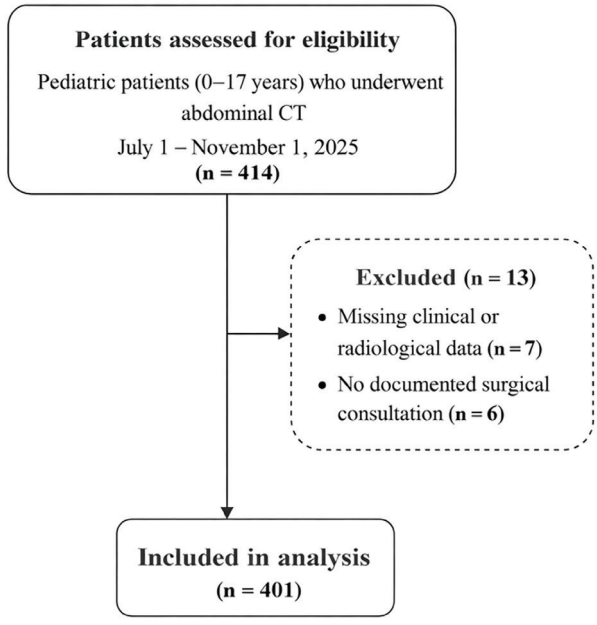

Figure 1. Flow diagram illustrating patient selection and inclusion process. Pediatric patients (0-17 years) undergoing abdominal CT between July 1 and November 1, 2025 were assessed for eligibility (n=414). Thirteen patients were excluded due to missing clinical or radiological data (n=7) or lack of documented surgical consultation (n=6). A total of 401 patients were included in the final analysis

CT: Computed tomography

Surgical intervention was not determined solely by CT findings; operative decisions were based on the integration of clinical presentation, physical examination findings, laboratory parameters, imaging results, and the judgment of the attending pediatric surgeon.

Incidental Findings

Abdominal CT findings that were not directly related to the presenting indication but were considered clinically relevant were recorded separately as incidental findings. These included, but were not limited to: incidental cysts (ovarian, renal, hepatic); congenital anatomical variations (horseshoe kidney, malrotation); incidental solid organ lesions; vascular anomalies; lymphadenopathy unrelated to the primary indication; and urolithiasis detected in patients presenting for non-urological complaints. Incidental findings were documented for subsequent clinical follow-up.

Outcome Definition

The primary outcome was the requirement for surgical intervention (yes/no), defined as operative management for confirmed or strongly suspected intra-abdominal pathology during the index hospitalization. Surgical decisions were made by the attending pediatric surgical team based on clinical, laboratory, and imaging findings, consistent with current evidence-based practice guidelines.

For suspected appendicitis, particular consideration was given to imaging findings, inflammatory markers, physical examination, and, when available, documented clinical scores such as the Alvarado score or the Pediatric Appendicitis Score.

In trauma cases, surgical intervention was indicated for hemodynamically unstable patients who were unresponsive to resuscitation or for those with evidence of hollow viscus injury, ongoing hemorrhage, or peritoneal signs. For adnexal pathology, surgical intervention was performed for suspected or confirmed ovarian torsion, for which timely diagnostic laparoscopy is indicated to preserve ovarian function; ovarian cysts were managed operatively only when associated with torsion, rupture, hemorrhage, or significant mass effect. For mass lesions, surgery was indicated when malignancy could not be excluded or when the lesion caused significant symptoms.

For intussusception, first-line management consisted of enema reduction; surgical intervention was warranted in patients who failed nonoperative management, were hemodynamically unstable despite resuscitation, had evidence of bowel compromise or ischemia, or had suspicion of a pathologic lead point. For gastrointestinal obstructive conditions, including ileus, surgical intervention was indicated in cases refractory to conservative management or with evidence of ischemia or perforation, consistent with the ACR Appropriateness Criteria for bowel obstruction management. For pancreatitis, surgical intervention was reserved for cases with infected necrosis, abdominal compartment syndrome refractory to conservative management, or perforated viscus. For gastrointestinal inflammatory conditions, including enteritis and colitis, surgical intervention was indicated when conservative management failed, peritoneal signs developed, or complications such as perforation or abscess formation were confirmed. For urolithiasis, surgical or endoscopic intervention was considered in cases of obstruction, superimposed urinary tract infection, intractable pain, or failure of spontaneous stone passage. In hepatobiliary pathology, intervention was indicated for biliary obstruction accompanied by clinical deterioration or failure of conservative management. Surgical intervention for gastrointestinal bleeding was indicated in cases of hemodynamic instability or ongoing hemorrhage unresponsive to endoscopic or conservative management. Mesenteric lymphadenitis and constipation were managed conservatively in all cases; surgical intervention was not indicated.

Non-specific CT findings were correlated with the clinical status, and surgical intervention was reserved for patients with progressive clinical deterioration or the development of peritoneal signs. In selected cases, surgical intervention may be considered despite negative imaging findings if progressive clinical deterioration or persistent peritoneal signs raise concern for significant intra-abdominal pathology.

Incidental findings did not independently constitute surgical indications, as detailed in the Incidental Findings section. Procedures included appendectomies, laparotomies, trauma-related interventions, and adnexal- or mass-related surgeries.

Statistical Analysis

All statistical analyses were performed using IBM SPSS Statistics for Windows, version 26.0 (IBM Corp., Armonk, NY, USA). Continuous variables were expressed as mean ± standard deviation or median (interquartile range), depending on data distribution, while categorical variables were presented as frequencies and percentages. Normality was assessed using the Shapiro-Wilk test.

Comparisons between independent groups were performed using the Student's t-test or Mann-Whitney U test, as appropriate. Categorical variables were analyzed using the chi-square test or Fisher's exact test.

Abdominal CT findings (CT^+/CT^-) and surgical intervention (yes/no) were defined as binary variables. The relationship between CT findings and the requirement for surgery was evaluated using 2 × 2 contingency tables.

No multivariable predictive modeling was performed, because the primary objective of this study was to

describe the spectrum of abdominal CT findings and to assess their association with surgical intervention rather than to develop a prediction model. Accordingly, model calibration, discrimination analyses (e.g., receiver operating characteristic curves), and internal validation procedures were not undertaken.

The association between CT findings and surgical intervention was evaluated using odds ratios (ORs) with 95% confidence intervals (CIs). In addition, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated to describe the relationship between CT findings and surgical intervention. A two-tailed p-value <0.05 was considered statistically significant.

Results

Among the 401 patients who underwent abdominal CT, 377 (94.0%) had at least one pathological finding (CT^+), whereas 24 (6.0%) had no pathological findings (CT^-). Demographic characteristics are summarized in Table 1. The mean age was 11.5±4.5 years, and 54.1% of patients were male.

Using surgical intervention as the reference outcome, the PPV was 30.0% and the NPV was 95.8%. Computed tomography positivity was significantly higher among females with pelvic/ovarian pathology (p<0.001), and these patients were older than those without pelvic pathology (13.27±3.42 vs. 10.64±4.11 years, p=0.002). The distribution of clinical indications is summarized in Table 2.

Suspected appendicitis, trauma, and pelvic/adnexal pathologies were the most common indications, whereas a heterogeneous "other" group accounted for 13.5% of cases. The most frequent CT findings were non-specific abnormalities, appendiceal pathology, trauma-related findings, and ovarian/adnexal pathology (Table 3; Figure 2).

Table 1. Demographic characteristics of the study population

Characteristic	Value
Total patients	401
Age, median (IQR)	12 (8-15)
Age, mean ± SD	11.5±4.5
Sex, n (%)	
Male	217 (54.1%)
Female	184 (45.9%)
Nationality, n (%)	
Turkish	337 (84.0%)
Non-Turkish	64 (16.0%)

Values are presented as mean ± standard deviation, median (interquartile range), or number (%)
SD: Standard deviation, IQR: Interquartile range

The relationship between clinical indications and CT-detected pathologies is illustrated in Figure 3. Acute abdominal pain was mainly associated with appendiceal pathology, trauma with trauma-related findings, and pelvic pain with ovarian/adnexal pathologies. Differences between ICD-10-coded preliminary diagnoses and CT-confirmed findings were observed, particularly for appendicitis, trauma, and ovarian and adnexal pathologies (Supplementary Table S1).

Table 2. Indications for abdominal CT (n=401)

Indication group	n	%
Suspected/confirmed appendicitis	91	22.7
Trauma	72	18.0
Pelvic/adnexal pathology	53	13.2
Gastrointestinal inflammation	31	7.7
Urolithiasis/urinary stone disease	25	6.2
Mesenteric lymphadenitis	21	5.2
Ileus/obstruction	15	3.7
Mass/neoplasm	9	2.2
Constipation	8	2.0
Hepatobiliary pathology	6	1.5
Gastrointestinal bleeding	5	1.2
Pancreatitis	4	1.0
Non-specific abdominal pain	7	1.7
Other	54	13.5
Total	**401**	**100.0**

Values are presented as number (%)
CT: Computed tomography

Table 3. Spectrum of CT findings and associated additional abnormalities

CT finding category	n (patients)	%	Additional findings, n (%)
Appendiceal pathology	87	21.7	18 (4.5)
Ovarian/adnexal pathology	39	9.7	10 (2.5)
Trauma-related findings	75	18.7	27 (6.7)
Gastrointestinal inflammatory/obstructive conditions	14	3.5	12 (3.0)
Urologic pathology	9	2.2	13 (3.2)
Mesenteric lymphadenitis	3	0.7	7 (1.7)
Neoplastic lesions	5	1.2	14 (3.4)
Other non-specific findings	169	42.1	15 (3.7)

Additional findings include incidental or secondary abnormalities detected on CT. Primary CT diagnostic categories are mutually exclusive. Additional findings are not mutually exclusive and were recorded independently of the primary CT diagnostic category; therefore, they could occur in patients from any primary category. Additional findings are presented as the number and percentage of the total study population (n=401)
CT: Computed tomography

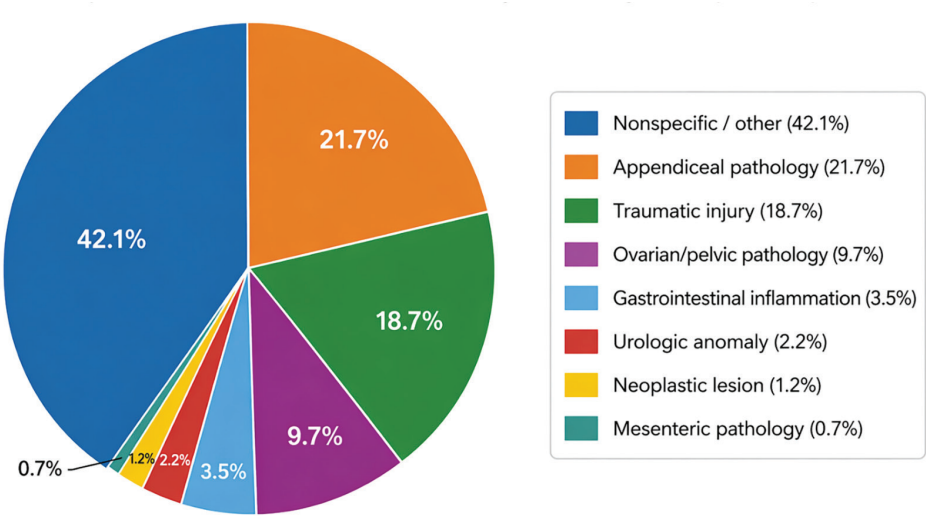

Figure 2. Proportional distribution of CT-detected diagnostic categories in pediatric patients (n=401). Non-specific/other findings (42.1%), appendiceal pathology (21.7%), and traumatic injury (18.7%) were the most frequently observed categories
CT: Computed tomography

Additional findings beyond the primary CT diagnostic categories were identified in 116 patients (28.9%) (Supplementary Table S2). Representative imaging findings are shown in Supplementary Figures S1 and S2. The distribution of indications for surgical consultation is presented in Figure 4. Acute appendicitis and trauma accounted for the majority of consultations.

Table 4 presents the association between CT findings and surgical intervention. Surgical intervention was performed in 113 (30.0%) CT-positive patients and in 1 (4.2%) CT-negative patient. Among the 24 CT-negative patients, one underwent surgical intervention because of progressive clinical deterioration and persistent peritoneal signs despite negative imaging findings. Computed tomography positivity was significantly associated with surgical intervention (p=0.007), with an OR of 9.84 (95% CI: 1.3-72.5), as illustrated in Figure 5.

Among the 75 trauma cases, 90.7% of injuries were blunt, and 5.3% of patients required surgical intervention. Detailed subgroup characteristics are presented in Supplementary Table S3.

A total of 20 mass lesions were identified in 13 patients (lesion-based analysis) (Supplementary Table S4). Nine patients underwent CT for a suspected mass or tumor, while lesions in four patients were detected incidentally.

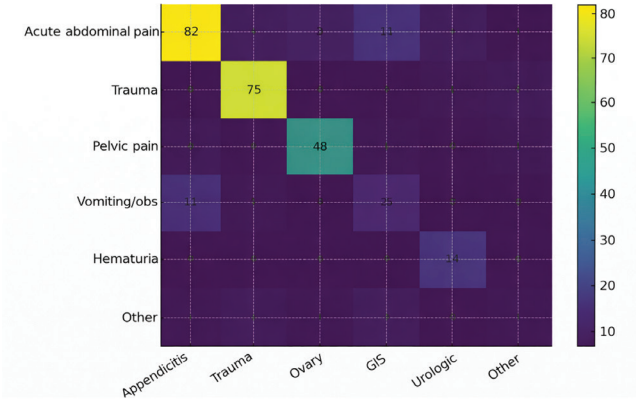

Figure 3. Heatmap illustrating the relationship between clinical indications and CT-detected diagnostic categories. Rows represent clinical indications, and columns represent CT diagnostic categories. Cell values indicate the number of patients
CT: Computed tomography

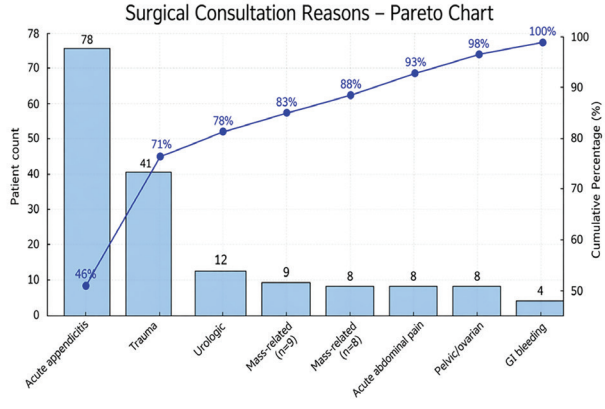

Figure 4. Distribution of indications for surgical consultation and cumulative contribution (Pareto chart). Bars represent the number of patients for each indication category, and the line represents the cumulative percentage
GI: Gastrointestinal

Table 4. Association between CT findings and surgical intervention					
	Surgery (–)	**Surgery (+)**	**Total**	**OR (95% CI)**	**p-value**
CT-negative	23	1	24	Reference	-
CT-positive	264	113	377	9.84 (1.3-72.5)	0.007
Total	287	114	401	-	-

Values are presented as number of patients. The association between CT findings and surgical intervention was evaluated using the chi-square test
OR: Odds ratio, CI: Confidence interval, CT: Computed tomography

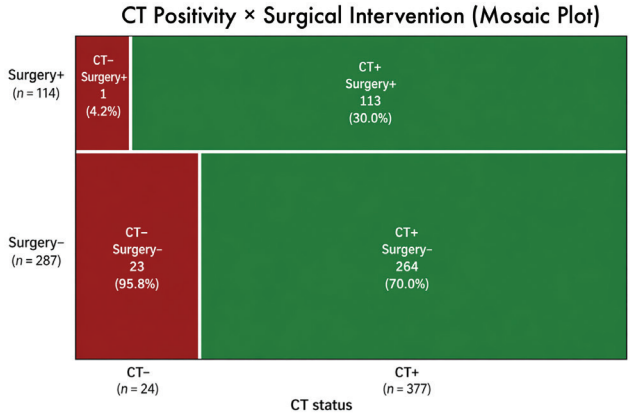

Figure 5. Association between CT positivity and surgical intervention (mosaic plot). The mosaic plot illustrates the relationship between CT findings (CT^+ vs. CT^-) and surgical intervention (yes/no). The size of each rectangle is proportional to the number of patients in each category
CT: Computed tomography

Discussion

In this tertiary pediatric surgical cohort, abdominal CT was most frequently performed for suspected appendicitis, trauma, and pelvic/adnexal pathology. The high CT positivity rate observed in our cohort (94.0%) suggests that abdominal CT was used selectively in patients with a relatively high pretest probability of significant pathology, rather than as a routine first-line imaging modality. This interpretation is supported by the distribution of CT indications observed in the study population. Furthermore, despite a high rate of pathological findings, surgical intervention was performed in only 28.4% of patients, indicating that CT findings frequently contributed to diagnostic clarification and to clinical assessment without necessarily leading to operative management.

The selective imaging pattern observed in our cohort differs from reports from general emergency settings, where CT utilization rates are substantially lower and patient populations are typically less enriched for surgical pathology (11-13). This finding likely reflects both the tertiary referral nature of our institution and the selective use of CT following initial clinical and ultrasonographic evaluation. The high CT positivity rate observed in the present study supports this interpretation and suggests that abdominal CT was primarily reserved for patients with persistent diagnostic uncertainty or concern for significant intra-abdominal pathology.

Our heatmap analysis demonstrated a strong alignment between clinical presentation and CT findings, particularly for acute abdominal pain and appendiceal pathology, trauma and trauma-related findings, and pelvic pain and ovarian/adnexal pathologies.

A substantial proportion of CT examinations were performed for heterogeneous "other" indications, reflecting the broad spectrum of clinical presentations encountered in pediatric emergency settings. This variability underscores the complexity of diagnostic decision-making and highlights the need for flexible, indication-based imaging strategies.

Subgroup analysis demonstrated significantly higher CT positivity among female patients with pelvic or ovarian pathologies ($p<0.001$). Patients with pelvic pathology were significantly older than those without pelvic pathology (13.27±3.42 vs. 10.64±4.11 years; $p=0.002$). These findings highlight the influence of age- and sex-related clinical characteristics on the spectrum of CT-detected pathology in pediatric patients.

These findings suggest the value of an indication-based imaging approach in pediatric emergency practice and should be interpreted within the context of a tertiary referral population.

Ovarian/adnexal pathology was identified in 39 patients (9.7%). These lesions may lead to clinically significant complications, including rupture, hemorrhage, and adnexal torsion (14). Together with the observed age- and sex-related differences, this finding highlights the important contribution of gynecologic causes to the spectrum of CT-detected pathology in adolescent females. This observation emphasizes the need to maintain a broad differential diagnosis when evaluating abdominal pain in older female patients.

Non-specific abdominal findings constituted the largest diagnostic category on abdominopelvic CT (42.1%). When interpreted in conjunction with clinical data, these findings may suggest conditions such as enteritis, colitis, or obstruction. However, this high proportion also reflects the frequency of non-specific CT findings in pediatric patients presenting with abdominal complaints and

underscores the broad differential diagnosis encountered in routine clinical practice. Similar observations have been reported in previous pediatric imaging studies, in which non-specific or clinically indeterminate CT findings were frequently encountered during the evaluation of abdominal symptoms (15).

Although non-specific findings were common, appendicitis remains the most frequent condition requiring emergency abdominal surgery in children (16). In our cohort, CT was frequently used in diagnostically uncertain cases and identified a broad spectrum of abdominal pathologies. Similar observations regarding the diagnostic utility of CT in abdominal pathology have been reported previously (4,17).

Computed tomography demonstrated a high NPV and a relatively low PPV (30.0%), indicating that many CT-positive findings were ultimately managed without surgical intervention. Notably, the vast majority of CT-negative patients did not require surgery (95.8%), suggesting that negative CT findings were generally associated with non-operative management in this cohort. Computed tomography positivity was significantly associated with surgical intervention (OR=9.84). The overall surgical intervention rate was 28.4%. Consistent with these findings, CT imaging has been reported to result in a change in clinical management in approximately one-third (34.8%) of children presenting with abdominal pain who underwent scanning (13). This pattern suggests that the clinical value of CT in a tertiary pediatric surgical setting extends beyond the identification of surgically treatable pathology and may also support confidence in non-operative management when clinically appropriate.

Suspected appendicitis was the most common indication for abdominal CT in the present cohort, and appendiceal pathology was among the most frequently identified CT findings, occurring in 21.7% of patients. In addition to appendiceal pathology, trauma was a major indication for abdominal CT in pediatric patients. Trauma-related findings were identified in 18.7% of all abdominal CT examinations. Previous studies on pediatric blunt abdominal trauma have reported that more than 80% of CT scans are negative, with intra-abdominal injury rates of approximately 17% (18), closely aligning with the distribution observed in the present cohort.

Blunt trauma accounted for the majority of trauma cases in this cohort (90.7%), consistent with the distribution typically observed in pediatric trauma populations. Trauma subgroup analysis demonstrated that the spleen, liver, and kidneys were the most commonly affected organs, reflecting the characteristic pattern of solid-organ injury in pediatric blunt abdominal trauma. Computed tomography enabled characterization of organ-specific injuries and provided additional information for clinical assessment. The low rate of surgical intervention (5.3%) supports the predominance of non-operative management in hemodynamically stable pediatric patients, which is considered the standard of care (19). However, negative FAST findings cannot reliably exclude intra-abdominal injury, and further imaging may be required in clinically suspicious cases.

In contrast to trauma-related findings, urinary system pathologies (2.2%) and tumoral lesions (1.2%) were relatively uncommon. The low frequency of tumoral lesions in our cohort likely reflects the selective use of abdominal CT for acute clinical presentations in a pediatric surgical setting, where inflammatory and traumatic conditions predominate. Among these lesions, the low rate of malignancy is consistent with previous reports indicating that many pediatric abdominal tumors are benign or exhibit low malignant potential (20).

Mass lesions were infrequently identified and, when present, were distributed across multiple anatomical locations, predominantly involving the ovaries, liver, and retroperitoneal regions. While some lesions were evaluated based on clinical suspicion, others were detected incidentally, highlighting the broader diagnostic scope of CT beyond the primary indication for imaging. The variability in lesion characteristics, including size, morphology, and associated features such as necrosis or lymphadenopathy, emphasizes the contribution of CT to lesion characterization and clinical assessment in pediatric patients. Although uncommon, these lesions represented a diagnostically heterogeneous group, highlighting the importance of comprehensive image interpretation in pediatric abdominal CT examinations (21,22).

A substantial proportion of ICD-10 entries were symptom- or mechanism-based codes, which contributed to a marked discordance between preliminary clinical diagnoses and findings detected by CT. Although suspected appendicitis was the most frequently recorded ICD-10 diagnosis (48.8%), appendiceal pathology was identified in only 21.7% of patients. This discrepancy suggests that ICD-10 coding in pediatric emergency settings may favor sensitivity over specificity, potentially leading to overestimation of certain surgical diagnoses.

The marked discrepancy between referral diagnoses and CT findings highlights the challenges of establishing a definitive diagnosis based solely on initial clinical assessment and supports the use of structured imaging pathways in selected patients with equivocal presentations.

In this context, cross-sectional imaging may contribute to refinement of the differential diagnosis when clinical findings are non-specific or overlapping. Similar findings have been reported in pediatric appendicitis, where CT and magnetic resonance imaging demonstrate high diagnostic accuracy in resolving diagnostic uncertainty (23).

Among specific diagnoses, mesenteric lymphadenitis was rare in our series (0.7%) and was generally managed conservatively, consistent with previous reports describing its typically benign clinical course (24).

In our study, incidental or secondary findings were present in 28.9% of patients, highlighting the broader diagnostic scope of CT beyond the primary clinical indication. Although most of these findings were not directly related to the acute presentation, they may still have clinical relevance. Their relatively high frequency underscores the importance of systematic image review and suggests that CT may occasionally detect additional findings that could influence subsequent patient evaluation; however, their impact on clinical management could not be assessed in the present study.

The need for surgical intervention in pediatric consultations was concentrated in a limited number of diagnostic categories, with acute appendicitis and trauma accounting for the majority of cases. This distribution was consistent with the pattern of CT indications observed in our cohort, where suspected appendicitis and trauma were among the most common reasons for abdominal CT examinations. Similar consultation patterns have been reported in pediatric emergency practice (25).

The higher number of urological consultations relative to CT-confirmed cases suggests a cautious clinical approach to suspected conditions and indicates that these pathologies are generally uncommon and often managed conservatively. This discrepancy between consultation patterns and CT-confirmed pathology highlights the complexity of diagnostic decision-making in pediatric abdominal presentations (13).

Overall, our findings demonstrate that abdominal CT was frequently used in cases with diagnostic uncertainty and that it identified a broad spectrum of pathologies in a tertiary pediatric surgical population. The high proportion of pathological findings, together with the relatively low rate of surgical intervention, reflects the selective use of CT and the heterogeneous nature of pediatric abdominal presentations.

Unlike most pediatric CT studies that focus on a single disease entity, this study evaluated abdominal CT utilization across diverse pediatric surgical presentations within a tertiary referral setting. The integration of clinical indications, CT findings, consultation patterns, and surgical outcomes provides a pragmatic overview of how abdominal CT contributes to everyday decision-making in pediatric surgical practice.

These findings are consistent with current recommendations supporting selective CT use following initial ultrasonographic evaluation in pediatric patients (26).

Study Limitations

This study has several limitations. First, variability in clinical decision-making regarding CT indications—particularly for heterogeneous presentations that required individualized clinical judgment—together with incomplete laboratory data and the lack of consistently documented clinical scoring systems for all patients may have introduced residual confounding. Although consecutive inclusion of all eligible CT examinations reduced sampling bias, selection bias related to the decision to perform CT examinations could not be fully eliminated because imaging decisions were made as part of routine clinical care rather than according to a prospectively standardized study protocol. Second, the single-center, retrospective design may limit the generalizability of the findings. Pathologic confirmation was available only in surgically treated cases, representing an inherent limitation of retrospective analyses. Additionally, variability in imaging practices, including externally performed USG and heterogeneity in CT acquisition parameters, may have affected the consistency of imaging interpretation. Limited access to long-term follow-up data further limits the ability to correlate findings with clinical outcomes. Because the primary objective of the study was descriptive rather than predictive, multivariable predictive modeling was not performed. Accordingly, the findings should not be used to derive clinical decision thresholds. Associations between CT findings and surgical intervention were not adjusted for potential confounders such as age, sex, or clinical indication and should therefore be interpreted descriptively rather than as evidence of an independent effect of CT findings on surgical decision-making.

Despite these limitations, this study has several strengths, including a large pediatric cohort from a tertiary surgical center and a comprehensive analysis of the relationship between clinical indications and radiologic findings. The use of heatmaps, proportional distribution analyses, Pareto charts, and mosaic plots enabled a detailed characterization of diagnostic patterns, imaging utilization, and surgically relevant findings across a broad spectrum of clinical presentations. Additionally, the inclusion of patients with diverse clinical indications and the structured imaging approach contributed to the representativeness of the cohort and enhanced the clinical applicability of the findings.

Conclusion

Abdominal CT scans in a tertiary pediatric surgical center were most commonly performed for suspected appendicitis, trauma, and presentations related to pelvic or adnexal pathology. Despite a high CT positivity rate,

the relatively low rate of surgical intervention indicates that CT serves as a complementary, not definitive, tool in surgical management. The high NPV observed in this study suggests that negative CT findings were generally associated with nonoperative management of clinically equivocal cases.

Ethics

Ethics Committee Approval: The study protocol was approved by the University of Health Sciences Türkiye, Gaziantep City Hospital Non-Interventional Clinical Research Ethics Committee (approval no: 348/2025, date: 17.12.2025).

Informed Consent: This study was conducted retrospectively using hospital data. The requirement for informed consent was waived by the ethics committee due to the retrospective nature of the study.

Authorship Contributions

Surgical and Medical Practices: S.B.S., Concept: S.B.S., Design: S.B.S., Data Collection or Processing: S.B.S., R.B.P., M.U., Analysis or Interpretation: S.B.S., R.B.P., M.U., Literature Search: S.B.S., Writing: S.B.S.

Conflict of Interest: No conflicts of interest were declared by the authors.

Financial Disclosure: This study received no financial support.

Supplementary Tables Link: https://d2v96fxpocvxx.cloudfront.net/68ab204c-182b-49da-b227-bc7efe058632/content-images/61e031b2-8073-4c99-a839-680419bea7fe.pdf

Supplementary Figures Link: https://d2v96fxpocvxx.cloudfront.net/68ab204c-182b-49da-b227-bc7efe058632/content-images/4a6ba4eb-450f-46d8-b969-6c55890b601a.pdf

References

1. Buel KL, Wilcox J, Mingo PT. Acute abdominal pain in children: evaluation and management. Am Fam Physician. 2024;110:621-31.
2. Michelson KA, Reeves SD, Grubenhoff JA, et al. Clinical features and preventability of delayed diagnosis of pediatric appendicitis. JAMA Netw Open. 2021;4:e2122248.
3. Expert Panel on Pediatric Imaging:; Koberlein GC, Trout AT, Rigsby CK, et al. ACR appropriateness criteria® Suspected Appendicitis-Child. J Am Coll Radiol. 2019;16:S252-63.
4. Baykara AS. Acute appendicitis in children: evaluation of the diagnostic efficacy of ultrasonography and computed tomography. Cureus. 2023;15:e43860.
5. El-Nakeep S, Tafti D, Pozun A. Intestinal ultrasonography: technique, indications, and clinical applications. [Updated 2026 Jun 19]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026 Jan-.
6. Li RT, Zhao Y, Zou XJ, et al. Overview of point-of-care ultrasound in diagnosing intestinal obstruction. World J Emerg Med. 2022;13:135-40.
7. Mohammad SA, Elsayed RA, Arafa NE, et al. Computed tomography of pediatric abdominal trauma: optimizing utilization and enhancing diagnostic interpretation. Pediatr Radiol. 2025;55:2018-36.
8. Blomerus S, Splinter TL, Gillis A, Buckley O, Turner H, McCabe A. Implementation of an acute abdominal pain diagnostic pathway in the emergency department. BMJ Open Qual. 2025;14:e003505.
9. Solomon DZ, Ayalew B, Dellie ST, Admasie D. Justification and optimization principles of ALARA in pediatric CT at a teaching hospital in Ethiopia. Ethiop J Health Sci. 2020;30:761-6.
10. Singer C, Saban M, Luxenburg O, et al. Computed tomography referral guidelines adherence in Europe: insights from a seven-country audit. Eur Radiol. 2025;35:1166-77.
11. Ahmad T, Khdair Ahmad F, Manson D. Diagnostic performance of a staged pathway for imaging acute appendicitis in children. Pediatr Emerg Care. 2021;37:e1197-201.
12. El-Gohary Y, Molina M, Chang J, et al. The use of computed tomography versus clinical acumen in diagnosing appendicitis in children: a two-institution international study. J Pediatr Surg. 2021;56:1356-61.
13. Khan HH, Subedi S, Kumar S, Lyons H. The pattern of CT scan use in the diagnosis of abdominal pain in children presenting to the emergency department of a tertiary community Hospital. Cureus. 2021;13:e19162.
14. Mobeen S, Apostol R. Ovarian Cyst. 2023 Jun 5. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026.
15. Ugalde IT, Yen K, Tatro G, et al. Incidental findings on computed tomography in children with blunt abdominal trauma. Ann Emerg Med. 2025;85:405-10.
16. Waseem M, Wang CF. Pediatric Appendicitis. 2025 Jun 17. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026.
17. Annu A, Garg UC, Sylonia SC. Evaluation of role of computed tomography in acute abdomen: a comprehensive clinical analysis. Int J Med All Body Health Res. 2025;6:128-33.
18. Otaibi BW, Khurshid MH, Hejazi O, et al. The abdomen does not lie, but the labs might: predictors of intra-abdominal injury on computed tomography imaging in pediatric blunt trauma patients. J Trauma Acute Care Surg. 2025;99:181-7.
19. Bašković M, Keretić D, Lacković M, Borić Krakar M, Pogorelić Z. The diagnosis and management of pediatric blunt abdominal trauma-a comprehensive review. Diagnostics (Basel). 2024;14:2257.
20. von Rohr R, Pachl M. Abdominal and pelvic tumours in children: diagnosis, management, and prognosis. Minim Invasive Pediatr Surg. 2025;43:510-9.
21. Jha SK, Brown C, Kang L, et al. Update on the role of imaging in staging of common pediatric abdominal tumors. Curr Probl Cancer. 2023;47:100969.

22. Gbadamosi H, Mensah YB, Appau AA, Renner LA. A spectrum of findings on computed tomography in paediatric abdominal and pelvic tumours in a Ghanaian teaching hospital. Ghana Med J. 2022;56:295-302.

23. Castro-Luna DI, Porras-Hernandez JD, Flores-Garcia JA, Dies-Suarez P, Servin-Martinez MF, Pierdant-Perez M. Contemporary ultrasound, computed tomography, or magnetic resonance imaging for acute appendicitis diagnosis in children and adolescents: systematic review and meta-analysis. Pediatr Radiol. 2025;55:1448-64.

24. Yukcu B, Yildirmak ZY, Ozel A, Genc DB. Clinical evaluation and outcomes of mesenteric lymphadenopathy in children: a cross-sectional analysis. Sisli Etfal Hastan Tip Bul. 2025;59:127-33.

25. Kalyoncu Aycenk A, Salci G. Appropriateness of pediatric surgical consultations from the emergency department: a 2-center study. Med Sci Monit. 2025;31:e951263.

26. American Academy of Pediatrics Committee on Pediatric Emergency Medicine, Section on Radiology; American College of Emergency Physicians Pediatric Emergency Medicine Committee; American College of Radiology; Marin JR, Lyons TW, Claudius I, et al. Optimizing advanced imaging of the pediatric patient in the emergency department: technical report. J Am Coll Radiol. 2024;21:e37-69.

Original Article

DOI: 10.4274/haseki.galenos.2026.17362
Med Bull Haseki 2026;64(4):304-314

Job Satisfaction, Burnout, and Mental Health Among Child Advocacy Center Professionals: A Cross-sectional Comparative Study

Fatmagul Aslan[1], Hacer Yasar Teke[2], Ender Atabay[3], Muslum Kul[3], Zehra Zerrin Erkol[4]

[1]*University of Health Sciences Türkiye, Antalya Training and Research Hospital, Clinic of Forensic Medicine, Antalya, Türkiye*
[2]*Ordu University Faculty of Medicine, Department of Forensic Medicine, Ordu, Türkiye*
[3]*Mersin City Training and Research Hospital, Clinic of Child and Adolescent Psychiatry, Mersin, Türkiye*
[4]*Bolu Abant Izzet Baysal University Faculty of Medicine, Department of Forensic Medicine, Bolu, Türkiye*

Abstract

Aim: Professionals working at child advocacy centers (CACs) perform their duties under demanding professional, social, ethical, and legal conditions. This study aimed to compare job satisfaction, burnout, depressive symptoms, and dimensions of general psychiatric symptoms among CAC employees with those of a control group.

Methods: This cross-sectional study included 30 healthcare professionals working as forensic interviewers at five CACs in Türkiye and 30 healthcare professionals from the same occupational groups without CAC experience. Participants completed a structured socio-demographic questionnaire, the Symptom Checklist-90-Revised, the Beck Depression Inventory, and the Maslach Burnout Inventory. Data were collected between December 1 and December 10, 2018.

Results: Compared with controls, the Children's Advocacy Center group had longer professional experience, lower job satisfaction, and higher depressive and overall psychiatric symptom scores, while burnout levels did not differ between groups.

Conclusion: Job satisfaction scores were found to be statistically significant predictors of depression scale scores, whereas age and length of service significantly affected burnout scale scores.

Keywords: Psychological tests, job satisfaction, health personnel, children's advocacy center, child welfare

Introduction

Burnout was first introduced into the scientific literature in the 1970s by Freudenberger and Richelson (1). Recently, burnout has attracted increasing attention across various disciplines and is particularly prevalent in occupations involving intensive interpersonal interaction. Healthcare professionals are consistently among the groups most affected by burnout, a condition emerging as a prolonged response to chronic emotional and interpersonal stressors (2).

According to the World Health Organization's ICD-11, burnout is defined as a syndrome resulting from chronic workplace stress that has not been successfully managed (3). Kristensen et al. (4) distinguished three dimensions of burnout: personal, work-related, and client-related. Research on employees in child protection units in the United States suggests that both work- and client-related burnout symptoms follow a curvilinear path over time, with job rotation providing a protective benefit during the first two years (5).

Burnout has adverse consequences for both employees and service recipients, being associated with chronic fatigue, sleep disturbances, impaired concentration, and psychological distress (6). Healthcare professionals

Corresponding Author: Prof. Dr. Hacer Yasar Teke, Ordu University Faculty of Medicine, Department of Forensic Medicine, Ordu, Türkiye
E-mail: hacer.hgulderen2004@gmail.com **ORCID:** orcid.org/0000-0003-2311-5145
Received: 13.05.2025 **Accepted:** 14.03.2026 **Publication Date:** 24.09.2026

Cite this article as: Aslan F, Yasar Teke H, Atabay E, Kul M, Erkol ZZ. Job satisfaction, burnout, and mental health among child advocacy center professionals: a cross-sectional comparative study. Med Bull Haseki. 2026;64(4):304-314

experiencing burnout are more prone to depression, substance use, poor physical health, and even suicidal ideation (7). Furthermore, studies highlighting the alarmingly high prevalence of burnout among healthcare workers emphasize the importance of evaluating both individual coping strategies and interventions aimed at improving working conditions (8).

Child advocacy centers (CACs) represent stressful and demanding work environments where professionals are in direct contact with children exposed to trauma (9). Heavy workloads, staff shortages, secondary traumatic stress, legal pressures, and the severe psychological issues of children and families place substantial strain on these professionals, potentially leading to burnout and increased turnover intention (10).

We hypothesized that professionals working in CACs would exhibit higher levels of mental health problems, including burnout and depression, than their peers. This study aimed to compare burnout levels, job satisfaction, and psychiatric symptoms between CAC professionals and a control group and to examine socio-demographic factors to identify potential risk and protective factors.

Materials and Methods

Compliance with Ethical Standards

The study was conducted in accordance with the principles of the Declaration of Helsinki. The study was approved by the University of Health Sciences Türkiye, Antalya Training and Research Hospital Ethics Committee (approval number: 20/6, date: 12.09.2019).

Study Design and Setting

This cross-sectional comparative study was conducted between December 1 and December 10, 2018, at five CACs located in five provinces in Türkiye. These centers provide multidisciplinary medico-legal and psychosocial services to children suspected of being sexually abused.

The study population consisted of healthcare professionals who were actively working as forensic interviewers at five CACs in Türkiye at the time of data collection and had been in their current positions for at least six months. Forensic interviewers at CACs included social workers, nurses, child development specialists, psychologists, and midwives.

The study population consisted of healthcare professionals working as forensic interviewers at CACs, while the control group comprised professionals from the same occupational backgrounds who worked in other hospital departments and had no previous experience with CACs (Figure 1).

A total of 30 participants were included in each group. Participants were contacted either by direct on-site visits or by e-mail. All participants were informed about the study, and only those who volunteered and provided written informed consent were enrolled.

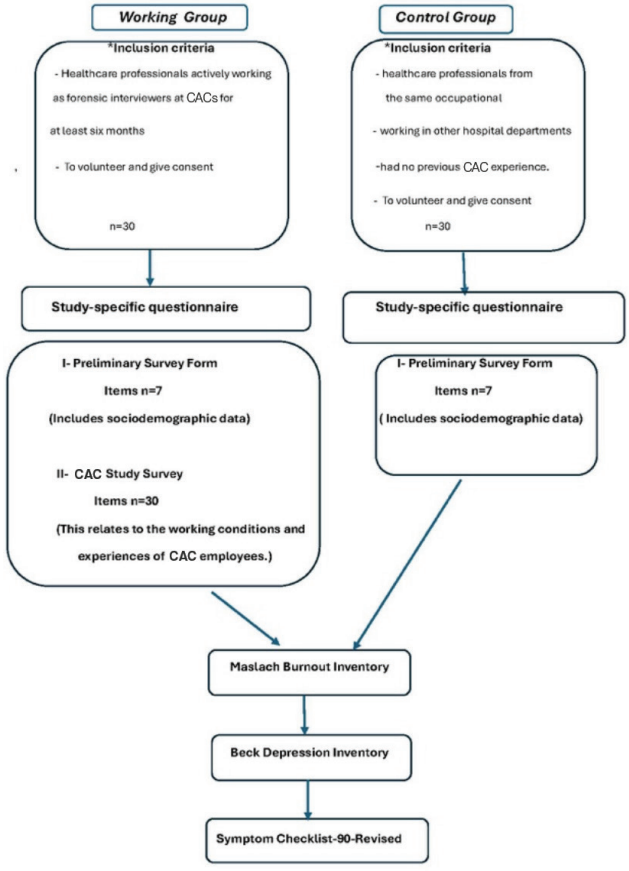

Figure 1. Design diagram of the study
CAC: Child advocacy center

Physicians working at CACs were not included in the study because of different job roles, additional professional responsibilities, distinct motivational factors (such as other clinical duties and financial incentives), and exposure to different work-related stressors.

Background Information on the CACs: in Türkiye, CACs have become increasingly widespread and operate on a 24-hour on-call basis. Each center is staffed by specially trained professionals, including a physician responsible for coordination and trained healthcare personnel such as social workers, psychologists, child development specialists, and nurses. In 2011, the first CAC in Türkiye opened in Ankara. By September 2019, the number of CACs had reached 44 (11).

For the study planned in 2018, the article published by Oh and Lee (12) in 2009 was used as a reference, and a power analysis was conducted assuming an effect size of 0.5. With an alpha error probability of 0.05 and a power of 0.80, the total sample size was calculated to be 34.

Measures

A study-specific questionnaire, the Maslach Burnout Inventory (MBI), the Beck Depression Inventory (BDI), and the Symptom Checklist-90-Revised (SCL-90-R) were administered to assess burnout levels among professionals working in CACs and to identify variables associated with burnout.

Study-Specific Questionnaire: It comprised two sections. The first section included nine items, derived from the literature, to obtain socio-demographic information, including age, sex, profession, duration of professional experience, duration of work in the current unit or center, self-rated job satisfaction and happiness scores, and membership in professional organizations.

The second section consisted of 30 items designed to evaluate experiences related to working at CACs and perceptions of the centers and of referred cases. All items were developed specifically for this study to enable CACs' staff to self-assess their working conditions and professional experiences. Responses were recorded using a three-point Likert-type scale. Because this section contained CACs-specific items, it was administered only to the study group.

The MBI consists of 22 items rated on a five-point Likert scale (six response options in the original version) and assesses burnout across three dimensions: emotional exhaustion (nine items), depersonalization (five items), and personal accomplishment (eight items). Unlike emotional exhaustion and depersonalization, the personal accomplishment subscale is reverse-scored. Higher scores on the emotional exhaustion and depersonalization subscales and lower scores on the personal accomplishment subscale indicate higher levels of burnout. No specific cut-off values are defined for the subscale scores (13,14).

Beck Depression Inventory: The BDI is a 21-item self-report scale rated on a four-point Likert scale and used to assess the presence and severity of depressive symptoms. Total scores range from 0 to 63, and a cut-off score of 17 is commonly used to indicate clinically relevant depressive symptomatology (15,16).

Symptom Checklist-90-Revised: The SCL-90-R is a 90-item self-report instrument developed to assess a broad range of psychological symptoms and distress. It can be used in both the general population and clinical settings. The scale gives scores for nine main symptom dimensions: somatization, obsessive-compulsive symptoms, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. It also gives a global severity index that shows the overall severity of the symptoms (17,18).

Procedure

After written informed consent was obtained, all participants completed the questionnaires and scales in a single session. Data collection was conducted over one week and completed on 10 December 2018. All data were collected anonymously, and no personally identifiable information was recorded.

Statistical Analysis

Data were analyzed using SPSS version 22.0 (Statistical Package for the Social Sciences). Categorical variables were compared using the chi-square test. The Shapiro-Wilk test was used to assess normality. For non-normally distributed continuous variables, the Mann-Whitney U test and Spearman's correlation analysis were employed. For normally distributed data, independent samples t-tests and one-way analysis of variance with Tukey post-hoc tests were used. Additionally, multiple linear regression analysis was performed to identify independent predictors of burnout and depression scores. Statistical significance was set at $p<0.05$.

Results

Socio-demographic Data and Job Satisfaction

A total of 60 healthcare professionals were included in the study: 30 working at five CACs and 30 without CAC experience. The CACs group was older (38.76±9.03) and had longer (15.06±10.41) professional experience ($p<0.05$). Participants' socio-demographic and professional characteristics are summarized in Table 1.

Job satisfaction scores were significantly lower in the CACs group compared with the control group ($t=-3.68$, $p<0.05$); professional association membership and non-professional hobbies did not differ between groups. Job satisfaction scores showed a significant negative correlation with BDI scores ($r=-0.4$, $p<0.05$), emotional exhaustion scores ($r=-0.44$, $p<0.05$), and SCL-90-R depression scores ($r=-0.31$, $p<0.05$) across all participants.

Evaluation of the Questionnaire Form

The CACs' specific questionnaire was completed only by the CACs' employees. Responses indicated high perceived importance of forensic and child psychiatric evaluations, high compatibility with colleagues, and a strong belief in the necessity and continuity of CACs. A considerable proportion of participants reported work-related stress and emphasized the negative impact of organizational deficiencies on their performance. The distribution of responses is shown in Table 2.

Responses to the questionnaire were analyzed by grouping CACs employees into psychologists, social workers, and other personnel. Multiple comparison tests

Table 1. Socio-demographic and professional data of the groups			
	Working group	**Control group**	**p-value**
Age [median (IQR)]	38 (14)	30 (13.25)	p<0.01*
Gender	22 (73%) female 8 (27%) male	11 (36%) female 19 (64%) male	p<0.01* χ^2=7.877**
Occupation groups	5 (16%) psychologist 4 (13%) child development specialists 11 (36%) social workers 8 (26%) nurses 1 (3%) midwife 1 (3%) secretary	12 (40%) psychologist 1 (3%) child development specialists 15 (50%) social workers 2 (7%) nurses	
Professional association membership	22 (73%) yes 8 (27%) no	19 (63%) yes 11 (37%) no	p=0.40 χ^2=0.69**
Seniority (average ± SD)	15.06±10.41	8.40±6.90	p<0.01* t=2.92
Professional satisfaction point [median (IQR)]	75 (25)	85 (10)	p<0.01*
Non-professional hobby assets	21 (70%) positive 9 (30%) negative	18 (60%) positive 12 (40%) negative	p=0.65 χ^2=0.41**

*Refers to the values that reach statistical difference between both groups
**Chi-square and t-tests were used for statistical analysis
IQR: Interquartile range, SD: Standard deviation

revealed statistically significant differences in four items. Healthcare personnel working outside the psychologist and social worker groups were significantly more likely both to believe that cases of child abuse are increasing (χ^2=11.74 , p<0.05) and to support the view that specially trained CACs personnel should assist physicians during forensic examinations (χ^2=6.34 , p<0.05).

Regarding the statement, "The most appropriate social decision for our cases is made by the responsible Public Prosecutor," social workers were significantly less likely to agree compared with other personnel (z=-2.4, p<0.05). Social workers were also more likely than other groups to support the opinion that "CAC staff should be renewed every three to four years" (z=-3.3, p<0.05).

Evaluation of Scale Scores

The median BDI score was significantly higher among CACs' personnel than in the control group (p<0.05), whereas no statistically significant difference was found between the groups for mean MBI scores (p>0.05). In addition, the overall symptom median score on the SCL-90-R test and the median scores of the somatization, obsessive-compulsive, interpersonal sensitivity, depression, anger-hostility, and paranoid ideation subscales were statistically significantly higher in CACs' employees than in controls (p<0.05). Comparisons of psychometric scale scores between groups are presented in Table 3.

In terms of socio-demographic variables, no significant effect on SCL-90-R scores was observed; however, the only parameter influencing the mean BDI score was the job satisfaction score. The group with the lowest job satisfaction had a significantly higher mean BDI score (F=7.85, p<0.05), and Tukey post-hoc analysis revealed a significant difference between the lowest and highest job satisfaction groups (p<0.05) (Table 4).

For the MBI, mean scores were affected by both professions and job satisfaction levels. Social workers had higher mean burnout scores than other professional groups (F=9.58, p<0.05). Similarly, job satisfaction scores influenced mean burnout scores in a manner comparable to that of the depression scale (F=5.89, p<0.05) (Table 5).

In the multiple regression model developed for the BDI, only the "job satisfaction score" emerged as a statistically significant predictor of total scale scores (p<0.05) (Table 6). Regarding the MBI, both age and length of professional experience had nearly equal, statistically significant impacts on burnout scores (p<0.05) (Table 7).

Discussion

Maslach et al. (13) emphasized that the primary determinants of burnout are the "areas of work life" and that burnout can be prevented only by systematically addressing and modifying these areas. These areas include workload, control, reward, community, fairness, and values. Time pressure also constitutes an important source of stress. Urgent medical examinations to prevent loss of forensic evidence, pressure related to detention or arrest decisions for alleged offenders, scheduling of forensic interviews and urgent placement of children into protective care represent major time-related stressors (19). Internal organizational factors, such as supervisor and colleague support, have been shown to reduce burnout (20). However, little is known about the impact of external organizational resources (e.g., therapists, advocates, police officers, and mental health professionals) on burnout.

Table 2. CAC employees' answers to the questions in the questionnaire			
Question	**Percentage values**		
	Yes	**Partially**	**No**
I can follow the latest legal and scientific regulations in my area of expertise	20%	66%	14%
I think I am compatible with my colleagues at CAC	83%	10%	7%
"Consent" should be obtained before all the procedures for children who are applied to us.	83%	10%	7%
The relevant public prosecutor makes the best social decision for our cases	17%	53%	30%
When the case comes, I can be stressed if I am a forensic interviewer on duty	17%	40%	43%
I can allocate enough time to cases when I am a forensic interviewer	50%	50%	-
I find it easier to be a family interviewer than a forensic interviewer.	54%	23%	23%
Meeting with the family is easier and less stressful than meeting the child.	33%	27%	40%
When the information about the future of the case reaches CAC, I will make a preliminary preparation if I am a forensic interviewer.	90%	7%	3%
After my judicial interviews, I generally think that I can reach sufficient judicial opinion.	60%	40%	-
Forensic examination is important	97%	3%	-
Child psychiatric report is important	97%	3%	-
It will be better for patients that the lawns are in the hospital	76%	14%	10%
CAC employees should be renewed every 3-4 years	40%	26%	34%
CAC employees should be supported financially	76%	21%	3%
Visiting workshops to other centers in terms of their professional competence should be organized for those who require social responsibility, such as CAC, and should be financed by the state.	86%	14%	-
I received psychiatric support while I was working at CAC	-	6%	94%
Working in lawns is psychologically	40%	60%	-
My work at CAC affected my perspective on life	57%	37%	6%
I like kids	93%	7%	-
My work at CAC influenced my value judgments about children's innocence	10%	30%	60%
My work at lawns affected my value judgments about adults	40%	50%	10%
It will be better to have a nurse trained in this area alongside our physicians who carry out the forensic examination.	90%	10%	-
The lack of a secretary among CAC employees adversely affects my job performance.	60%	20%	20%
I think cases of child abuse are increasing	86%	14%	-
I think that child abuse reporting is increasing	100%	-	-
I think that the most correct social decision can be taken after the social review report in cases of child abuse.	63%	34%	3%
When I forward the reports of the case to the relevant prosecutor, I believe that this issue will be the most effective decision.	53%	40%	7%
I think that the forms of such centers determined by law should not be changed for many years and should be independent of politics unless necessary.	87%	13%	-
I think physical conditions are effective in my work performance	84%	8%	8%
It contains descriptive statistical data CAC: Child advocacy centers			

Some studies indicate that child abuse professionals who establish positive relationships with external professionals experience less frustration, feel less isolated, and are better able to fulfill their responsibilities, thereby potentially reducing burnout (21). There is broad consensus in the literature that professionals who work in the field of child neglect and abuse are at high-risk of burnout. However, more detailed studies are needed to better understand the specific challenges encountered in this field (22,23).

Despite the specialized nature of this field, reduced job satisfaction and burnout have been identified as the main determinants of turnover among professionals working in CAC-like settings (2). Role conflict, role ambiguity and role overload, safety concerns, insufficient supervisory support, secondary traumatic experiences, and compassion fatigue have been reported as major risk factors affecting job satisfaction and burnout (2). Stressful working conditions may further increase vulnerability to depressive symptoms

Table 3. Examination of the scale scores of the groups			
	CAC	**Control**	**p-value**
Beck Depression Inventory	8.83±5.36	5.50±5.92	p<0.05*
Maslach Burnout Inventory	11.86±6.45	10.56±6.45	p=0.43
Maslach personal success	10.13±4.70	10.90±5.76	p=0.57
Maslach desensitization	4 (2.5)	3 (6)	p=0.67**
SCL somatization	0.58 (0.58)	0.25 (0.29)	p<0.01*,**
SCL OCD	0.60 (0.80)	0.40 (0.58)	p<0.05*,**
SCL interpersonal	0.66 (0.63)	0.33 (0.50)	p<0.01*,**
SCL depression	0.69 (0.58)	0.26 (0.39)	p<0.01*,**
SCL anxiety	0.40 (0.60)	0.25 (0.40)	p=0.16**
SCL anger hostility	0.33 (0.67)	0.16 (0.41)	p<0.05*,**
SCL phobic anxiety	0.14 (0.28)	0.14 (0.42)	p=0.94**
SCL paranoid thought	0.83 (0.76)	0.16 (0.54)	p<0.01*,**
SCL psychotism	0.10 (0.30)	0.10 (0.30)	p=0.33**
SCL additional scale	0.71 (0.72)	0.49 (0.85)	p=0.08**
SCL overall symptom average	0.54 (0.50)	0.32 (0.45)	p<0.01*,**

*Refers to the values that reach statistical difference between both groups
**Mann-Whitney U and t-tests were used for statistical analysis
CAC: Child advocacy center, SCL: Symptom checklist, OCD: Obsessive-compulsive disorder

Table 4. The effect of professional satisfaction point on Beck Depression Inventory score				
Professional satisfaction point	**n**	**Mean ± SD**	**F**	**p-value**
0-70	15	11.80±5.77*	9.58	p<0.01
71-89	27	5.48±1.05		
90 and above	18	4.00±3.94*		

*This indicates groups that show a statistically significant difference in the Tukey test, one-way ANOVA was used for statistical analysis
ANOVA: Analysis of variance, SD: Standard deviation

and substantially contribute to reduced job satisfaction and burnout (22). Burnout and job satisfaction are also reported to influence each other reciprocally (22), while individual characteristics and personal resilience may play either protective or vulnerability-enhancing roles (10,22).

Depression is a multifactorial disorder resulting from interactions among genetic, developmental, and environmental factors. Burnout and reduced job satisfaction have been shown to be associated with depression and other mental disorders, with chronic psychosocial stress playing a central role in both conditions (24).

Individuals who are dissatisfied with their work and who cannot find a supportive organizational climate are more likely to experience marked declines in morale and motivation (25). Lower job satisfaction has been reported among professionals conducting abuse-related interviews in the literature (23,26). In a study conducted among social workers employed in child protection services in the United Kingdom, role ambiguity and job satisfaction were identified as predictors of stress, while job satisfaction was suggested to be influenced by both perceived stress levels and role conflict (27). Consistent with the literature, the present study found that job satisfaction scores from the preliminary questionnaire were statistically significantly lower in CAC employees than in the control group. One possible explanation for this finding is longer professional experience among healthcare workers in the CAC group compared with controls. Although longer professional experience may increase competence and expertise, it may also lead to reduced motivational resources over time (28). Although no CAC-specific studies are available, findings regarding the relationship between age and job satisfaction in the literature are contradictory (29). In our study, no statistically significant difference in job satisfaction was found among age groups (p>0.05).

Moral distress and secondary traumatic stress may contribute to this situation, and workplace peer relationships should be considered. A study conducted among social workers working in child protection services in the United Kingdom reported that participants associated workplace friendships with a continuously increasing sense of well-being, higher job satisfaction, reduced stress, and a

Table 5. The effect of socio-demographic data on Maslach Burnout Inventory score				
	n	**Mean ± SD**	**F**	**p-value**
Age				
30 years and under	20	11.80±6.96	0.40	0.66**
31-40 years	21	11.66±6.62		
41 years and older	19	10.10±5.81		
Seniority				
1-5	18	10.38±5.92	2.30	0.08**
6-10	16	13.00±7.75		
11-15	10	7.20±4.23		
16 years and older	16	12.87±5.84		
Occupation groups				
Psychologist	17	9.00±6.68*	5.51	p<0.01**
Social workers	17	14.15±5.95*		
Others	26	8.94±5.26*		
Professional satisfaction point				
0-70	15	15.20±1.60*	5.89	p<0.01**
71-89	27	11.11±5.97		
90 and above	18	8.05±5.71*		

*This indicates groups that show a statistically significant difference in the Tukey test
**One-way ANOVA was used for statistical analysis
ANOVA: Analysis of variance, SD: Standard deviation

Table 6. Multiple linear regression analysis for predictors of Beck Depression Inventory score							
	ß	**Std. error**	**Standardized coefficients ß**	**t**	**p-value**	**95% confidence interval ß**	
						Lower bound	**Upper bound**
Constant	17.14	0.7.2		2.3	0.02*	2.70	31.58
Age	0.09	0.20	0.14	0.49	0.62	-0.30	0.49
Gender	0.57	1.55	0.04	0.37	0.71	-2.54	3.69
Seniority	-0.03	0.18	-0.05	-0.18	0.85	-0.40	0.33
Professional satisfaction point	-0.16	0.05	-0.41	-3.26	p<0.01*	0.26	-0.06

*Refers to statistically significant value, multiple Linear regression analysis was used for statistical analysis

Table 7. Regression models developed for Maslach Burnout Inventory							
	ß	**Std. error**	**Standardized coefficients ß**	**t**	**p-value**	**95% Confidence interval ß**	
						Lower bound	**Upper bound**
Constant	39.34	8.08		4.86	p<0.01*	23.13	55.54
Age	-0.50	0.21	-0.67	-2.36	0.02*	-0.92	-0.07
Gender	-1.54	1.65	-0.12	-0.93	0.35	-4.86	1.77
Seniority	0.52	0.20	-0.76	2.59	0.01*	-0.11	0.92
Occupation groups	-3.90	2.16	-0.44	-1.80	0.07	-8.24	0.44
Professional satisfaction point	-0.06	0.10	-0.14	-0.61	0.54	-.0.28	-0.15

*Refers to statistically significant value, multiple Linear regression analysis was used for statistical analysis

greater intention to remain in their roles (30). Furthermore, several studies have reported that individual factors such as higher educational level, specific work environments, and self-efficacy are positively correlated with job satisfaction (31). The literature also includes studies emphasizing the importance of strengthening employee well-being as a strong predictor of job satisfaction and positive work attitudes among social workers (32,33).

In Türkiye, the lack of sufficiently developed regulations concerning CAC and the absence of a formal directive defining the division of duties among staff may contribute to decreased job satisfaction, since such legal uncertainties can limit the legal empowerment of professionals working under such conditions. Finally, previous studies have highlighted that job satisfaction plays an important role in maintaining psychological well-being and functional mental health (26).

In the present study, internal organizational factors other than colleague compatibility (93%) were not examined. Among external organizational factors, only perceptions regarding mental health professionals and legal professionals were assessed. Participants reported a very high level of belief in the importance of child psychiatry reports (97%), which may indirectly reflect trust in positive collaboration with these professionals. Healthcare personnel outside the psychologist and social worker groups were statistically more likely to believe that cases of child abuse are increasing and to support the view that specially trained CAC personnel should assist physicians during forensic examinations. However, compared with other personnel, social workers were more likely to support renewing CAC staff every three to four years while being less likely to endorse the view that the best social decisions for children should be made by public prosecutors. We consider that the difference between social workers (who are familiar with the field and receive specialized training prior to working at CACs and other personnel may be related to differences in training. Employing personnel who are not psychologists or social workers and who lack field-specific training may undermine workplace harmony, perceived social support, and the ability to cope with secondary traumatization.

No significant difference in burnout levels was observed between the study and control groups. This finding appears to be largely inconsistent with the existing literature, as higher burnout levels among professionals working with child abuse cases have been consistently reported (22). However, this result may be partly explained by the fact that during the study period CACs were established in relatively newer buildings compared with many hospitals (the oldest CAC had been in operation for only 7-8 years) and that the control group also consisted of healthcare professionals. Burnout syndrome among healthcare workers has been reported to be primarily characterized by frustration and depersonalization (34). In the regression model developed in our study, age and professional seniority were identified as the two most important parameters influencing the burnout scale. Although not statistically significant, a decrease in the mean scores of the MBI was observed as age increased, which is consistent with other studies conducted among healthcare workers in Türkiye (35,36).

While most studies in the literature emphasize the change in burnout over time, they propose different perspectives regarding the trajectory of this change (37). According to the accumulation model, a continuous positive correlation develops between professional seniority and burnout over time. In contrast, the "onset effect" model suggests that burnout increases as employees adapt to their positions but decreases and returns to baseline levels once they become accustomed to the role (38,39). These components are known to develop over time, and the mean duration of employment at CACs among participants in our study was 3.50±1.88 years. In another study, the peak period for workplace burnout was reported to be between 12 and 24 months (5). Considering that the average duration of employment in our study was well above this period and that change is more common than stability, the timing of the measurement point may not have been optimal.

It is also well known that unfavorable physical working conditions contribute to burnout (40). Moreover, social support mitigates workplace stressors such as high workload, exposure to trauma, and limited resources that may lead to burnout among child protection professionals (41). The level and nature of social support may vary across countries depending on administrative structures, organizational frameworks, and cultural and social contexts. In the literature, there is a study providing preliminary evidence that social workers' participation in support programs—which encompass psychological protective factors such as mindfulness, acceptance, attention regulation/decentering, self-compassion, non-attachment, non-avoidance, and non-aversion—has the capacity to improve anxiety and rumination, thereby reducing stress, emotional exhaustion, and depersonalization (42). The fact that the newly structured CACs in Türkiye are newer buildings than those of other health institutions, have internal physical structures more suitable for the personal needs of health personnel, provide social support between individuals and supervisors, and offer facilitated opportunities for health personnel to participate in both in-service training and other educational events, such as congresses, may have contributed to the lack of difference in burnout scores between the two groups.

In this study, when socio-demographic characteristics were evaluated, social workers had significantly higher MBI scores than other professional groups. Studies focusing on social workers have identified workload and conflict between professional and family responsibilities as primary contributors to emotional exhaustion (43). In addition, lower professional satisfaction scores among healthcare personnel were statistically associated with higher mean scores on the MBI.

A study conducted in South Korea among child protection professionals suggested that depressive mood, identified as the second most common risk factor associated with intention to leave the job, reduces job satisfaction, increases emotional exhaustion, and diminishes the sense of personal accomplishment (22). Consistent with these findings, our study demonstrated a statistically significant difference in BDI scores between the study and control groups. Furthermore, healthcare personnel with lower professional satisfaction scores showed statistically significant increases in mean scale scores, a pattern similar to that observed in the MBI. Considering that the present study was conducted in newly established CACs, burnout levels may not have been fully captured; however, in line with the literature, the findings clearly indicate the presence of depressive symptoms and highlight the significant impact of professional satisfaction on both depression and burnout. Moreover, examination of the regression models demonstrating the negative linear relationships between job satisfaction scores and both the depression and burnout scales revealed that the contribution of the depression scale was nearly comparable to that of the burnout scale (Table 6).

The study found that median SCL-90 symptom scores, particularly the subscale medians for somatization, obsessive-compulsive symptoms, interpersonal sensitivity, depression, anger-hostility, and paranoid ideation, were significantly higher among CACs' employees than in the control group. Higher levels of interpersonal sensitivity may reflect the search for social support, which has been described in the literature as a factor that helps reduce workplace stress (41). The presence of findings of somatization may also represent a natural psychological manifestation of stress. In professionals working in highly demanding fields such as child neglect and abuse, the emotional processes described as compassion fatigue may contribute to anger as a result of grief related to perceptions of the world's safety and children's protectability (44,45).

Study Limitations

This study has several limitations. Because some questions were personal, response bias, particularly social desirability bias, may have occurred. Furthermore, participants were assessed cross-sectionally, and no longitudinal follow-up was performed; therefore, psychiatric symptom levels and scale scores may have been influenced by recent life events. Individual perceptions and attributions may also change over time depending on personal and professional circumstances. Repeating the study with the same sample at a later stage may provide additional insights. In addition, a substantial proportion of participants were not interviewed face-to-face; therefore, the impact of this on the honesty of responses remains unknown. Another limitation is that psychiatric symptomatology was assessed using self-report scales rather than structured or semi-structured psychiatric interviews. The relatively modest sample size and potential residual confounding related to baseline differences between the groups may also have influenced the findings. Despite these limitations, this study provides valuable preliminary evidence and contributes to the limited literature on healthcare professionals working in child monitoring centerss caring for children exposed to sexual abuse.

Conclusion

The present findings indicate that professionals working in CACs experience depressive mood states, a higher burden of psychiatric symptoms, and lower levels of job satisfaction. Given that reduced job satisfaction appears to be associated with both burnout and depression, further comprehensive quantitative and qualitative research addressing factors related to job demands, working system characteristics, and conditions, as well as individual-level determinants that may influence job satisfaction scores, may contribute to the literature. In addition, psychoeducational workshops focusing on stress management, mental health, self-care, personal strengths, and skill development and structured in-service training programs aimed at supporting professional development and advancement within the field may be encouraged.

Ethics

Ethics Committee Approval: Ethical approval for the study was obtained from the University of Health Sciences Türkiye, Antalya Training and Research Hospital Ethics Committee (approval number: 20/6, date: 12.09.2019).

Informed Consent: All participants were informed about the study, and only those who volunteered and provided written informed consent were enrolled.

Footnotes

Authorship Contributions

Concept: F.A., E.A., Z.Z.E., Design: F.A., M.K., Z.Z.E., Data Collection or Processing: F.A., E.A., M.K., Analysis or Interpretation: F.A., Literature Search: F.A., H.Y.T., Writing: F.A., H.Y.T.

Conflict of Interest: The authors declared that there were no conflicts of interest.

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

References

1. Freudenberger HJ, Richelson G. Burn-out : the high cost of high achievement. Garden City, N.Y.: Anchor Press; 1980.
2. Sprang G, Craig C, Clark J. Secondary traumatic stress and burnout in child welfare workers: a comparative analysis of occupational distress across professional groups. Child Welfare. 2011;90:149-68.
3. World Health Organization. Burn-out an "occupational phenomenon": International Classification of Diseases. 2019.
4. Kristensen TS, Borritz M, Villadsen E, Christensen KB. The Copenhagen Burnout Inventory: a new tool for the assessment of burnout. Work and Stress. 2005;19:192-207.
5. Radey M, Wilke DJ. Early career burnout trajectories among child protective services workers. Work and Stress. 2025;39:416-33.
6. De la Fuente-Solana EI, Gómez-Urquiza JL, Cañadas GR, Albendín-García L, Ortega-Campos E, Cañadas-De la Fuente GA. Burnout and its relationship with personality factors in oncology nurses. Eur J Oncol Nurs. 2017;30:91-6.
7. Williams ES, Rathert C, Buttigieg SC. The personal and professional consequences of physician burnout: a systematic review of the literature. Med Care Res Rev. 2020;77:371-86.
8. Grumbach K, Willard- Grace R. Health worker burnout and moral injury: drives, effects and remedies. Annual Review of Public Health. 2025;46:447-65.
9. Aslan F, Erkol ZZ. Türkiye'de Çocuk İzlem Merkezleri ve Antalya Çocuk İzlem Merkezi Deneyimi. Nobel Tıp Kitabevi; İstanbul 2023.
10. Russ E, Lonne B, Lynch D. Increasing child protection workforce retention through promoting a relational-reflective framework for resilience. Child Abuse Negl. 2020;110:104245.
11. Orhan G, Ulukol B, Canbaz H. The model of child advocacy center and problems observed in this model. Kriz Dergisi. 2019;27:178-86.
12. Oh S-H, Lee M. Examining the psycometric properties of the Maslach Burnout Inventory with a sample of child protective service workers in Korea. Children and Youth Services Review. 2009;31:206-10.
13. Maslach C, Jackson SE, Leiter MP. The Maslach Burnout Inventory manual. In: Zalaquett CP, Wood RJ, editor. Evaluating stress: a book of resources. The Scarecrow Press; 1997. p. 191-218.
14. Çapri B. Tükenmişlik ölçeğinin Türkçe uyarlaması: geçerlik ve güvenirlik çalışması. Mersin Univ Egit Fak Derg. 2006;2:62-77.
15. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. Arch Gen Psychiatry. 1961;4:561-71.
16. Hisli Sahin N. Beck depresyon envanteri'nin geçerliği üzerine bir çalışma. Turk Psikol Derg. 1988;22:118-26.
17. Derogatis LR, Unger R. Symptom checklist-90-revised. In: The Corsini Encyclopedia of Psychology (eds I.B. Weiner and W.E. Craighead), 2010.
18. Kılıç M. Belirti Tarama Listesi (SCL-90-R)'nin geçerlik ve güvenirliği. TPCGJ. 1991;1:45-52.
19. Bakker AB, Demerouti E. Job demands-resources theory: taking stock and looking forward. J Occup Health Psychol. 2017;22:273-85.
20. He AS, Phillips JD, Lizano EL, Rienks S, Leake R. Examining internal and external job resources in child welfare: protecting against caseworker burnout. Child Abuse Negl. 2018;81:48-59.
21. Garstka TA, Lieberman A, Biggs J, Thompson B, Levy MM. Barriers to cross-systems collaboration in child welfare, education, and the courts: supporting educational well-being of youth in care through systems change. J Public Child Welf. 2014;8:190-211.
22. Chung Y, Choo H. A structural path to job satisfaction, burnout and intent to leave among child protection workers: a South Korean study. Children and Youth Services Review. 2019;100:304-12.
23. Bonach K, Heckert A. Predictors of secondary traumatic stress among children's advocacy center forensic interviewers. J Child Sex Abus. 2012;21:295-314.
24. Bakusic J, Schaufeli W, Claes S, Godderis L. Stress, burnout and depression: a systematic review on DNA methylation mechanisms. J Psychosom Res. 2017;92:34-44.
25. Smith RJ, Clark SJ. Does job resource loss reduce burnout and job exit for social workers? Children and Youth Services Review. 2011;33:1950-9.
26. Chung Y, Chun JS. Workplace stress and job satisfaction among child protective service workers in South Korea: focusing on the buffering effects of protective factors. Child Youth Serv Rev. 2015;57:134-40.
27. Stanley S, Murphy C, Brougham R. How does work support influence the relationshipbetween stress and job satisfaction? (A Quantitative study of Social Workers in the UK). Journal of Social Work Practice. 2026. Available from:s URL:https://www.tandfonline.com/doi/pdf/10.1080/02650533.2025.2609657?casa_token=oriy2iJi0PUAAAAA:1qR9LxCQT_lHvdEeWejFgGH8KN2aPn-F0-OL45GZ1Rjo3ZRodeacU-ddLBBUyY-FmXTV8wenyHPkpw
28. Ardıç K, Polatcı S. Tükenmişlik sendromu: akademisyenler üzerinde bir uygulama. Gazi Univ IIBF Derg. 2008;10:69-96.
29. Dobrow Riza S, Ganzach Y, Liu Y. Time and job satisfaction: a longitudinal study of the differential roles of age and tenure. Journal of Management. 2018;44:2558-79.
30. Murphy C, Bedford A. What is the nature and impact of workplace friends within child protection social work? Implications for emotional support, professional learning, job satisfaction and worker wellbeing. British Journal of Social Work. 2025;53:1121-40.

31. Barth RP, Lloyd EC, Christ SL, Chapman MV, Dickinson NS. Child welfare worker characteristics and job satisfaction: a national study. Soc Work. 2008;53:199-209.

32. Alboroto R, Garza T, McNaughtan J. Enhancing caseworker job satisfaction through empowerment: an innovative tool for employee retention. Child Maltreat. 2025;30:553-64.

33. Brimhall KC, Lizano EL, Mor Barak ME. The mediating role of inclusion: a longitudinal study of the effects of leader–member exchange and diversity climate on job satisfaction and intention to leave among child welfare workers. Children and Youth Services Review. 2014;40:79-88.

34. Bai M, Jing T, Li S, Zhang Z. Unravelling critical burnout syndrome in Chinese hospitals a network analysis of the Maslach Burnout Inventory Human services survey. Sci Rep. 2025;15:9870.

35. Murat M, Köse S, Savaşer S. Determination of stress, depression and burnout levels of front-line nurses during the COVID-19 pandemic. Int J Ment Health Nurs. 2021;30:533-43.

36. Arslanoğlu A, İbrahimoğlu Ö, Mersin S, Ergül S, Yavuz AE. Burnout in healthcare workers: the effect of stress of conscience. J Eval Clin Pract. 2025;31:e14148.

37. Mäkikangas A, Kinnunen U, Feldt T, Schaufeli W. The longitudinal development of employee well-being: a systematic review. Work & Stress.2016;30:46-70.

38. Maslach C, Jackson SE. The measurement of experienced burnout†. J Organiz Behav. 1981;2:99-113.

39. Frese M, Zapf D. Methodological issues in the study of work stress: objective vs. subjective measurement of work stress and the question of longitudinal studies. In Cooper C.L, Payne R. (Eds.), Causes, coping, and consequents of stress at work. Wiley.1988:375-411.

40. Oliveira AM, Silva MT, Galvão TF, Lopes LC. The relationship between job satisfaction, burnout syndrome and depressive symptoms: an analysis of professionals in a teaching hospital in Brazil. Medicine (Baltimore). 2018;97:e13364.

41. Orock AN, Navratil P, Rasnayake S, Gakuba T-0."They stay, I stay, they quit, I quit": Mediating role of Social support on occupational risk and burnout among child protection Socail workers in Seychelles. European Journal of Social Work. 2026;1-13. Available from: URL:https://www.tandfo nline.com/doi/full/10.10 80/136 91457.2025.26 08838?sc roll=top& needAccess=true (Accessed December 23, 2026)

42. Maddock A. Examining potential psyschological protective and risk factors for stress and burnout in social workers. Clin Soc Work J. 2025;53:140-55.

43. Chaves-Montero A, Blanco-Miguel P, Ríos-Vizcaíno B. Analysis of the predictors and consequential factors of emotional exhaustion among social workers: a systematic review. Healthcare (Basel). 2025;13:552.

44. Albaek AU, Binder PE, Milde AM. Plunging into a dark sea of emotions: professionals' emotional experiences addressing child abuse in interviews with children. Qual Health Res. 2020;30:1212-24.

45. Cocker F, Joss N. Compassion fatigue among healthcare, emergency and community service workers: a systematic review. Int J Environ Res Public Health. 2016;13:618.

Original Article

DOI: 10.4274/haseki.galenos.2026.75436
Med Bull Haseki 2026;64(4):315-323

Frequency of Obesity, Elevated Blood Pressure, Anemia, and Micronutrient Deficiencies in a School-linked Adolescent Screening Cohort

Halil Ugur Hatipoglu, Elif Ozcan, Ozan Hayzaran, Muhammet Bayrak, Derya Buyukkayhan

University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Pediatrics, Istanbul, Türkiye

Abstract

Aim: School-linked adolescent screening may identify nutritional, hematologic, micronutrient, and blood pressure (BP) abnormalities in students outside regular preventive follow-up. This study evaluated the frequency of predefined, screening-detected abnormalities among adolescents who completed a school-linked clinical and laboratory screening pathway.

Methods: This retrospective cross-sectional study included 287 adolescents who attended pediatric outpatient screening after school-based outreach conducted between November 21 and December 6, 2023. Frequencies were reported with Wilson 95% confidence intervals (CIs). Multivariable logistic regression was used to assess factors associated with elevated or hypertensive-range BP.

Results: The sample included 204 girls and 83 boys; the median age was 14.0 years. Overweight/obesity was identified in 32.1% (95% CI, 26.9-37.7); elevated/hypertensive-range BP in 19.5% (95% CI, 15.3-24.5); anemia in 17.1% (95% CI, 13.1-21.9); ferritin <15 ng/mL in 24.7% (95% CI, 20.0-30.2); vitamin B12 <200 pg/mL in 18.8% (95% CI, 14.6-23.8); and folate <4 ng/mL in 15.1% (95% CI, 11.3-19.8). Boys [adjusted odds ratio (OR), 2.38; 95% CI, 1.17-4.83] and body mass index standard deviation score (aOR, 3.03; 95% CI, 2.17-4.25) were independently associated with elevated/hypertensive-range BP.

Conclusion: School-linked screening identified excess weight, abnormal BP screening findings, anemia, and micronutrient deficiencies. Such programs should integrate cardiometabolic assessment with sex-sensitive hematologic and micronutrient evaluation.

Keywords: Adolescent, school health services, mass screening, pediatric obesity, blood pressure, anemia, micronutrients

Introduction

Adolescence is a critical developmental period during which nutritional status, adiposity, blood pressure (BP) patterns, and micronutrient reserves may influence future cardiometabolic and hematologic health (1). Excess weight in childhood and adolescence remains a major public health concern, with recent global estimates showing a sustained and increasing burden of overweight and obesity among children and adolescents (2). Elevated BP during childhood and adolescence is also clinically important because pediatric hypertension predicts later cardiovascular risk, and recent systematic and scoping reviews have highlighted the global and European burden of pediatric hypertension, the higher risk among children and adolescents with overweight or obesity, and the relevance of school-based BP screening (3-5). At the same time, anemia, iron deficiency, and vitamin B12 insufficiency may occur during school age and adolescence and may remain clinically silent unless actively screened (6). Therefore, school-linked health screening may provide an opportunity to identify adolescents who do not routinely attend preventive healthcare visits (4).

In Türkiye, national surveillance data and previous primary care-based screening data indicate that excess weight, BP abnormalities, and anemia remain relevant health issues among school-aged children and adolescents

Corresponding Author: Lec, Halil Ugur Hatipoglu, MD, University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Pediatrics, Istanbul, Türkiye
E-mail: halilugur.hatipoglu@sbu.edu.tr **ORCID:** orcid.org/0000-0002-7393-677X
Received: 04.06.2026 **Accepted:** 23.08.2026 **Publication Date:** 24.09.2026

Cite this article as: Hatipoglu HU, Ozcan E, Hayzaran O, Bayrak M, Buyukkayhan D. Frequency of obesity, elevated blood pressure, anemia, and micronutrient deficiencies in a school-linked adolescent screening cohort. Med Bull Haseki. 2026;64(4):315-323

(7,8). School-linked screening activities may be organized through collaboration between education authorities and healthcare institutions, with students being informed at school and referred for clinical evaluation when appropriate. However, Turkish data evaluating anthropometric status, BP measurements, anemia, iron status, vitamin B12, and folate together in adolescent school-linked screening settings are limited. In addition, because laboratory testing is not always completed for all students reached during outreach activities, studies based on completed clinical and laboratory screening pathways should clearly distinguish analytic cohort frequencies from population-level prevalence estimates.

We hypothesized that a school-linked adolescent screening pathway would identify a substantial burden of excess weight, elevated/hypertensive-range BP, anemia, and micronutrient deficiencies, with sex- and body mass index (BMI)-related differences in selected abnormalities. Therefore, this study aimed to determine the frequency of predefined screening-detected abnormalities among adolescents who completed a school-linked clinical and laboratory screening pathway and to explore their associations with sex and BMI categories. By defining these patterns, the study may contribute to the development of more targeted adolescent screening and follow-up strategies that integrate cardiometabolic, hematologic, and micronutrient assessments.

Materials and Methods

Compliance with Ethical Standards

The study was approved by an University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital Institutional Non-Interventional Scientific Research Ethics Committee (approval no: 222-2026, date: 30.04.2026). Institutional administrative permission had also been obtained before the retrospective analysis. The study was conducted in accordance with the principles of the Declaration of Helsinki. The requirement for informed consent was waived because the study was retrospective and utilized anonymized data collected from routine screening.

Study Design and Setting

This retrospective cross-sectional study used records from a school-linked adolescent health screening pathway coordinated by a training and research hospital in collaboration with a district directorate of national education. The school-based outreach process and subsequent pediatric outpatient screening were conducted between November 21, 2023, and December 6, 2023. The participant flow is shown in Figure 1.

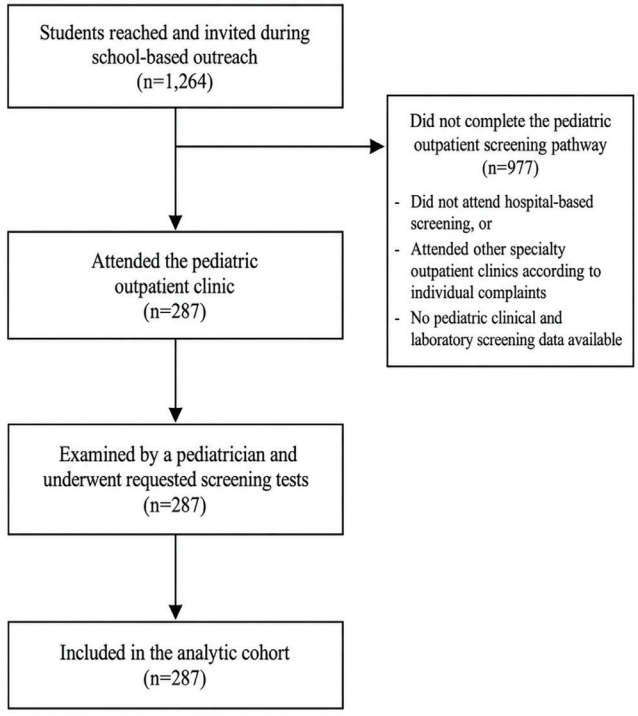

Figure 1. Participant flow of the school-linked adolescent screening cohort
A total of 1,264 students were reached and invited during school-based outreach. Of these, 287 attended the pediatric outpatient clinic, were examined by a pediatrician, underwent requested screening tests, and were included in the analytic cohort. Students who did not attend pediatric outpatient screening or who attended other specialty outpatient clinics were not included because pediatric clinical and laboratory screening data were not available

Participants, Inclusion Criteria, and Exclusion Criteria

During the school-based outreach phase, 1,264 adolescents were reached and informed. Students who had neither a known chronic disease nor regular follow-up at a healthcare center were invited to attend pediatric outpatient screening. Students with a known chronic disease or who were receiving regular healthcare follow-up were not targeted for this screening pathway.

The analytic cohort included adolescents who attended the pediatric outpatient clinic after school-based outreach and for whom anthropometric and BP measurements and relevant laboratory data were available. Adolescents with a known chronic disease were excluded. Data on medication use, vitamin or iron supplementation, and menstrual history were not systematically available. No pediatric clinical or laboratory screening data were available for students who did not attend the pediatric outpatient

clinic or who attended other specialty outpatient clinics for individual complaints. Therefore, the present analysis was restricted to 287 adolescents who completed the clinical and laboratory screening pathway.

Sample Size

Assuming a finite population of 23,400 high school students in the district, an expected prevalence of 50%, a 5% absolute precision, a 95% confidence level, and a design effect of 1, the minimum sample size required to estimate prevalence in the target population was calculated to be 378 using OpenEpi version 3. The final analytic sample was smaller than this target; therefore, frequency estimates were interpreted as findings from the analytic cohort rather than precise population-level prevalence estimates. Accordingly, this study should be considered an exploratory descriptive analysis of adolescents who completed the clinical and laboratory screening pathway. The reported estimates represent analytic cohort frequencies and should not be interpreted as population prevalence estimates.

Outcomes

The primary outcome was the frequency of predefined screening-detected abnormalities in the analytic cohort, including underweight, overweight, obesity, waist circumference ≥90th percentile, elevated/hypertensive-range BP, anemia, low ferritin concentration, transferrin saturation <16%, iron deficiency anemia, vitamin B12 deficiency, borderline/insufficient vitamin B12 status, and folate deficiency. Secondary outcomes were differences in these abnormalities by sex and BMI category and the independent associations of age, sex, and BMI standard deviation score (SDS) with elevated or hypertensive-range BP.

Measurements and Definitions

Age, sex, body weight, height, BMI, BMI percentile, BMI SDS, waist circumference, waist circumference percentile, mid-upper arm circumference (MUAC), systolic BP, diastolic BP, hemoglobin, serum iron, total iron-binding capacity, ferritin, vitamin B12, and folate were obtained from the hospital information system. Transferrin saturation was calculated as serum iron divided by total iron-binding capacity and then multiplied by 100.

Body mass index-based nutritional status was classified according to age- and sex-specific BMI percentile categories as underweight (<5th percentile), normal weight (5th to <85th percentile), overweight (85th to <95th percentile), and obesity (≥95th percentile) (9).

Mid-upper arm circumference was evaluated as a complementary anthropometric screening indicator and classified according to age- and sex-specific MUAC-for-age categories (10). Waist circumference at or above the 90th percentile was considered high.

Systolic and diastolic BP were measured by the same clinician using an oscillometric device with an appropriately sized cuff after 10 minutes of seated rest. A single blood pressure measurement was obtained during the screening encounter; repeated measurements were not averaged. Blood pressure categories were defined according to the 2017 American Academy of Pediatrics thresholds for adolescents aged 13 years or older (11). For this age group, normal BP was defined as systolic BP <120 mmHg and diastolic BP <80 mmHg; elevated BP as systolic BP 120-129 mmHg and diastolic BP <80 mmHg; stage 1-range BP as systolic BP 130-139 mmHg or diastolic BP 80-89 mmHg; and stage 2-range BP as systolic BP ≥140 mmHg or diastolic BP ≥90 mmHg. Because BP was measured during a single screening encounter, elevated values were interpreted as screening-detected elevated BP or as hypertensive-range BP, rather than confirmed hypertension.

Anemia was defined according to age- and sex-specific World Health Organization hemoglobin thresholds: hemoglobin <12 g/dL for adolescents aged 12-14 years and girls aged 15 years or older, and hemoglobin <13 g/dL for boys aged 15 years or older (12). Low ferritin concentration was defined as serum ferritin <15 ng/mL (13). Transferrin saturation <16% was also reported as a marker of low iron availability. Iron deficiency anemia was defined as the coexistence of anemia and a serum ferritin level <15 ng/mL. Vitamin B12 deficiency was defined as <200 pg/mL, and Vitamin B12 levels of 200-299 pg/mL were reported as borderline low. Vitamin B12 levels <300 pg/mL have also been reported to capture both deficient and borderline low values. Folate deficiency was primarily defined as <4 ng/mL, with <3 ng/mL also reported as a stricter threshold (14).

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 25.0 (IBM Corp., Armonk, NY, USA). Distributions of continuous variables were assessed using the Shapiro-Wilk test and visual inspection of normal probability plots. Normally distributed continuous variables were presented as mean ± standard deviation, whereas non-normally distributed variables were presented as median interquartile range (IQR). Categorical variables were summarized as numbers and percentages.

The frequencies of predefined screening-detected abnormalities were reported with 95% confidence intervals (CIs) calculated using the Wilson method. Comparisons between categorical variables were performed using Pearson's chi-square test or Fisher's exact test, as appropriate, based on expected cell counts. Statistically significant p-values were shown in bold in the tables.

Multivariable logistic regression analysis was performed to identify factors independently associated with elevated/hypertensive-range BP. Age, sex, and BMI SDS were

selected a priori based on clinical relevance and to avoid overfitting, considering the number of adolescents with elevated/hypertensive-range BP. Results were reported as adjusted odds ratios with 95% CIs. Model fit and performance were evaluated using the Hosmer-Lemeshow goodness-of-fit test, Nagelkerke R^2, classification accuracy, and the c-statistic/area under the receiver operating characteristic curve. A two-sided p-value <0.05 was considered statistically significant.

Results

A total of 1,264 students were reached and invited during the school-based outreach phase. Among them, 287 adolescents attended the pediatric outpatient clinic, were examined by a pediatrician, underwent the requested screening tests, and constituted the analytic cohort (Figure 1). Students who did not attend pediatric outpatient screening or attended other specialty outpatient clinics were excluded because pediatric clinical and laboratory screening data were unavailable.

The median age of the analytic cohort was 14.0 years (IQR, 14.0-14.0), and 204 participants (71.1%) were girls. Continuous clinical and laboratory characteristics are shown in Table 1, and categorical characteristics in Table 2. Overall, 21 adolescents (7.3%) were underweight, 42 (14.6%) were overweight, and 50 (17.4%) were obese. Thus, 92 adolescents (32.1%) were overweight or obese.

Table 1. Continuous clinical and laboratory characteristics of the analytic cohort

Variable	N	Summary statistic
Age, years	287	14.0 (14.0-14.0)
Height, cm	287	158.0 (155.0-164.0)
Weight, kg	287	55.0 (46.9-65.0)
BMI, kg/m²	287	21.45 (19.18-25.00)
BMI percentile	287	64.43 (29.46-90.82)
BMI SDS	287	0.35±1.33
Waist circumference, cm	287	70.0 (64.0-78.5)
Waist circumference percentile	287	48.0 (18.0-80.0)
MUAC, cm	285	24.0 (22.0-27.0)
Systolic BP, mmHg	287	110.0 (100.0-116.0)
Diastolic BP, mmHg	287	60.0 (60.0-72.0)
Hemoglobin, g/dL	281	13.1 (12.3-14.1)
Ferritin, ng/mL	275	25.4 (15.0-40.3)
Vitamin B12, pg/mL	272	268.5 (218.0-329.5)
Folate, ng/mL	272	5.9 (4.5-7.1)
Transferrin saturation, %	280	19.0 (12.9-26.7)

Values are presented as median (interquartile range) for non-normally distributed variables and mean ± standard deviation for normally distributed variables
BMI: Body mass index, BP: Blood pressure, MUAC: Mid-upper arm circumference, SDS: Standard deviation score

Waist circumference ≥90th percentile was identified in 49 adolescents (17.1%). Normal BP was recorded in 231 adolescents (80.5%), elevated BP in 15 (5.2%), stage-1 range BP in 38 (13.2%), and stage-2 range BP in 3 (1.0%).

The frequencies of predefined screening-detected abnormalities are presented with 95% confidence intervals in Table 3. Elevated or hypertensive-range BP was detected in 56 adolescents (19.5%; 95% CI, 15.3-24.5), including 41 adolescents (14.3%; 95% CI, 10.7-18.8) who had hypertensive-range BP. Anemia was detected in 48 of 281 adolescents (17.1%; 95% CI, 13.1-21.9); low ferritin concentration was detected in 68 of 275 adolescents (24.7%; 95% CI, 20.0-30.2); transferrin saturation <16% was detected in 104 of 280 adolescents (37.1%; 95% CI, 31.7-42.9); and iron deficiency anemia was detected in 30 of 274 adolescents (10.9%; 95% CI, 7.8-15.2). Vitamin B12 deficiency, defined as vitamin B12 <200 pg/mL, was present in 51 of 272 adolescents (18.8%; 95% CI, 14.6-23.8). Vitamin B12 <300 pg/mL, representing a broader category including both deficient and borderline low values, was present in 174 of 272 adolescents (64.0%; 95% CI, 58.1-69.4). Folate <4 ng/mL was detected in 41 of 272 adolescents (15.1%; 95% CI, 11.3-19.8).

Table 2. Categorical characteristics of the analytic cohort

Category	n (%)
Sex	
Girls	204 (71.1)
Boys	83 (28.9)
BMI category	
Underweight	21 (7.3)
Normal weight	174 (60.6)
Overweight	42 (14.6)
Obese	50 (17.4)
MUAC category	
Severe malnutrition	5 (1.7)
Moderate malnutrition	21 (7.3)
At risk of underweight	87 (30.3)
Normal nutritional status	165 (57.5)
At risk of overweight	7 (2.4)
Overweight	2 (0.7)
Obese	0 (0.0)
Blood pressure category	
Normal blood pressure	231 (80.5)
Elevated blood pressure	15 (5.2)
Stage 1-range blood pressure	38 (13.2)
Stage 2-range blood pressure	3 (1.0)

Values are presented as n (%). Percentages were calculated using the analytic cohort denominator (n=287). Blood pressure categories were based on a single screening measurement and do not indicate confirmed hypertension.
BMI: Body mass index, MUAC: Mid-upper arm circumference

Sex-based comparisons are shown in Table 4. Girls had higher frequencies of anemia, low ferritin concentrations, and iron deficiency anemia than boys. In contrast, boys had higher frequencies of waist circumference ≥90th percentile, elevated/hypertensive-range BP, and hypertensive-range BP. Vitamin B12 and folate abnormalities did not differ significantly by sex.

Table 3. Frequency of predefined screening-detected abnormalities

Finding	n/N (%)	95% CI
Adiposity and anthropometry		
Underweight	21/287 (7.3)	4.8-10.9
Overweight	42/287 (14.6)	11.0-19.2
Obesity	50/287 (17.4)	13.5-22.2
Overweight or obesity	92/287 (32.1)	26.9-37.7
Waist circumference ≥90th percentile	49/287 (17.1)	13.2-21.9
Blood pressure		
Elevated/hypertensive-range BP	56/287 (19.5)	15.3-24.5
Hypertensive-range BP, stage 1 or 2	41/287 (14.3)	10.7-18.8
Hematologic and micronutrient findings		
Anemia	48/281 (17.1)	13.1-21.9
Ferritin <15 ng/mL	68/275 (24.7)	20.0-30.2
Transferrin saturation <16%	104/280 (37.1)	31.7-42.9
Iron deficiency anemia	30/274 (10.9)	7.8-15.2
Vitamin B12 <200 pg/mL	51/272 (18.8)	14.6-23.8
Vitamin B12 200-299 pg/mL	123/272 (45.2)	39.4-51.2
Vitamin B12 <300 pg/mL	174/272 (64.0)	58.1-69.4
Folate <4 ng/mL	41/272 (15.1)	11.3-19.8
Folate <3 ng/mL	14/272 (5.1)	3.1-8.5

Values are presented as n/N (%). Confidence intervals were calculated using the Wilson method
BP: Blood pressure, CI: Confidence interval

Body mass index category-based comparisons are shown in Table 5. Waist circumference ≥90th percentile, elevated/hypertensive-range BP, and hypertensive-range BP differed significantly across BMI categories, with the highest frequencies observed among adolescents with obesity. Hematologic and micronutrient abnormalities did not differ significantly across BMI categories.

In multivariable logistic regression analysis, boys had higher odds of elevated/hypertensive-range BP than girls (aOR, 2.38; 95% CI, 1.17-4.83; p=0.016). BMI SDS was independently associated with elevated or hypertensive-range BP (aOR, 3.03 per 1-SD increase; 95% CI, 2.17-4.25; p<0.001), whereas age was not significantly associated with this outcome (aOR, 1.06; 95% CI, 0.61-1.86; p=0.835) (Table 6). The overall model was statistically significant (χ^2=66.562, df=3, p<0.001), with Nagelkerke R^2=0.330. The Hosmer-Lemeshow goodness-of-fit test did not indicate poor fit (χ^2=15.185, df=8, p=0.056). Overall classification accuracy was 85.7%, and the c-statistic/AUC was 0.817 (95% CI, 0.749-0.885; p<0.001).

Discussion

In this school-linked adolescent screening cohort, excess weight, elevated or hypertensive-range BP, anemia, low ferritin concentration, abnormal vitamin B12 levels, and folate deficiency were frequently detected. Approximately one-third of the analytic cohort had overweight or obesity, and nearly one-fifth had elevated/hypertensive-range BP. Boys had higher frequencies of abdominal obesity and abnormal BP screening findings, whereas girls had higher frequencies of anemia, low ferritin concentrations, and iron-deficiency anemia. In multivariable analysis, male sex and BMI SDS were independently associated with elevated or hypertensive-range BP. These findings suggest that school-linked screening pathways may identify clinically relevant cardiometabolic and hematologic abnormalities in

Table 4. Comparison of screening-detected abnormalities according to sex

Finding	Girls n/N (%)	Boys n/N (%)	p-value*
Anemia	46/200 (23.0)	2/81 (2.5)	**<0.001**
Ferritin <15 ng/mL	58/196 (29.6)	10/79 (12.7)	**0.003**
Transferrin saturation <16%	78/200 (39.0)	26/80 (32.5)	0.309
Iron deficiency anemia	28/195 (14.4)	2/79 (2.5)	**0.005**
Vitamin B12 <200 pg/mL	36/195 (18.5)	15/77 (19.5)	0.846
Vitamin B12 <300 pg/mL	121/195 (62.1)	53/77 (68.8)	0.294
Folate <4 ng/mL	29/195 (14.9)	12/77 (15.6)	0.882
Waist circumference ≥90th percentile	25/204 (12.3)	24/83 (28.9)	**<0.001**
Elevated/hypertensive-range BP	33/204 (16.2)	23/83 (27.7)	**0.025**
Hypertensive-range BP, stage 1 or 2	19/204 (9.3)	22/83 (26.5)	**<0.001**

Values are presented as n/N (%). *Pearson's chi-square test. Statistically significant p-values are shown in bold
BP: Blood pressure

Table 5. Comparison of screening-detected abnormalities according to BMI category					
Finding	**Underweight n/N (%)**	**Normal weight n/N (%)**	**Overweight n/N (%)**	**Obese n/N (%)**	**p-value**
Anemia	5/20 (25.0)	31/169 (18.3)	6/42 (14.3)	6/50 (12.0)	0.536†
Ferritin <15 ng/mL	7/21 (33.3)	46/168 (27.4)	7/42 (16.7)	8/44 (18.2)	0.267*
Transferrin saturation <16%	11/21 (52.4)	62/168 (36.9)	14/42 (33.3)	17/49 (34.7)	0.479*
Iron deficiency anemia	4/20 (20.0)	21/168 (12.5)	2/42 (4.8)	3/44 (6.8)	0.209†
Vitamin B12 <200 pg/mL	2/19 (10.5)	29/161 (18.0)	9/42 (21.4)	11/50 (22.0)	0.718†
Vitamin B12 <300 pg/mL	10/19 (52.6)	101/161 (62.7)	26/42 (61.9)	37/50 (74.0)	0.330*
Folate <4 ng/mL	4/19 (21.1)	22/161 (13.7)	8/42 (19.0)	7/50 (14.0)	0.656†
Waist circumference ≥90th percentile	0/21 (0.0)	6/174 (3.4)	7/42 (16.7)	36/50 (72.0)	**<0.001†**
Elevated/hypertensive-range BP	0/21 (0.0)	14/174 (8.0)	12/42 (28.6)	30/50 (60.0)	**<0.001†**
Hypertensive-range BP, stage 1 or 2	0/21 (0.0)	9/174 (5.2)	9/42 (21.4)	23/50 (46.0)	**<0.001†**

Values are presented as n/N (%).*Pearson's chi-square test; †Fisher's exact test. Statistically significant p-values are shown in bold
BMI: Body mass index, BP: Blood pressure

Table 6. Multivariable logistic regression analysis for elevated/hypertensive-range blood pressure			
Variable	**aOR**	**95% CI**	**p-value**
Age, per 1-year increase	1.06	0.61-1.86	0.835
Boys vs. girls	2.38	1.17-4.83	**0.016**
BMI SDS, per 1-SD increase	3.03	2.17-4.25	**<0.001**

The dependent variable was elevated/hypertensive-range blood pressure. The model included age, sex, and BMI SDS. Statistically significant p-values are shown in bold
aOR: Adjusted odds ratio, BMI: Body mass index, CI: Confidence interval, SDS: Standard deviation score

adolescents who would otherwise remain outside routine preventive follow-up.

The proportion of adolescents with overweight or obesity in the present analytic cohort was notable but should be interpreted cautiously because the cohort consisted of students who completed the pediatric outpatient screening pathway, rather than all students who were reached during school outreach. Nevertheless, the finding is consistent with the increasing burden of adolescent overweight and obesity reported in recent global estimates and with national data from Türkiye indicating that excess weight remains an important public health issue in school-aged children (2,7). Previous primary care-based data from Türkiye also support the relevance of obesity and related risk factors in children and adolescents (8). In the present study, waist circumference ≥90th percentile was particularly frequent among adolescents with obesity, indicating that excess adiposity was not limited to BMI-defined obesity but also included central adiposity, which is clinically relevant for cardiometabolic risk assessment.

Blood pressure findings are among the most clinically actionable outcomes of this study. Elevated or hypertensive-range BP was detected in 19.5% of adolescents; hypertensive-range BP was present in 14.3%. These values should not be interpreted as the prevalence of confirmed hypertension because BP was measured during a single screening encounter. However, they indicate a substantial need for repeated BP measurements and clinical follow-up. Recent systematic reviews have emphasized the global and European burden of high BP and hypertension in children and adolescents, as well as the association between excess weight and abnormal BP (3,5). In addition, a recent scoping review of school-based BP screening highlighted the heterogeneity of screening methods and the importance of clear referral and diagnostic pathways after elevated BP is detected (4). Our finding that BMI SDS was independently associated with elevated or hypertensive-range BP supports the need to integrate BP assessment into adolescent obesity screening pathways rather than relying on anthropometric measures alone.

The sex differences in BP findings also deserve attention. Boys had higher frequencies of elevated/hypertensive-range BP than girls, and male sex remained independently associated with elevated/hypertensive-range BP after adjustment for age and BMI SDS. This suggests that sex-related differences in adolescent BP screening findings may not be fully explained by adiposity alone. Because pubertal stage, physical activity, dietary sodium intake, family history, and repeated BP measurements were not available, the mechanisms underlying this difference could not be evaluated. Therefore, the observed association should be interpreted as a screening signal requiring follow-up rather than evidence of established sex-specific hypertension risk in this cohort.

Hematologic and iron-related abnormalities were also common. Anemia, low ferritin concentration, transferrin saturation <16%, and iron deficiency anemia were detected in 17.1%, 24.7%, 37.1%, and 10.9% of

adolescents, respectively. These findings are compatible with previous reports showing that iron-related abnormalities remain relevant among school-age children and adolescents and may be identified through screening-based approaches (6,15,16). Girls had markedly higher frequencies of anemia, low ferritin concentrations, and iron-deficiency anemia than boys. Menstrual blood loss, dietary iron intake, and sex-specific nutritional patterns may contribute to this difference; however, menstrual history and detailed dietary data were not systematically available in the present study. Therefore, the sex difference should be discussed as a clinically plausible finding rather than attributed to a single mechanism.

Findings for vitamin B12 and folate indicate that micronutrient assessment may provide additional information beyond anthropometry and BP measurements. Vitamin B12 concentrations <200 pg/mL and <300 pg/mL were detected in 18.8% and 64.0% of adolescents, respectively. The latter threshold includes both deficient and borderline low values and should therefore be interpreted as a broader screening indicator rather than a definitive indicator of biochemical deficiency. Previous pediatric and adolescent studies have also shown that vitamin B12 insufficiency may be frequent, although reported frequencies vary according to population characteristics and biochemical thresholds (6,17,18). Folate concentrations were detected in 15.1% of adolescents, whereas concentrations <3 ng/mL were less frequent. Because serum folate interpretation in otherwise healthy children and adolescents may be influenced by testing context and threshold selection, folate findings should be interpreted as screening signals requiring clinical correlation rather than as standalone diagnostic conclusions (14,19,20).

An important finding was that the frequency of hematologic and micronutrient abnormalities did not differ significantly across BMI categories, whereas the frequency of waist circumference and BP abnormalities increased markedly with higher BMI categories. This pattern suggests that BMI-based targeting alone may be insufficient for detecting anemia, low ferritin concentration, vitamin B12 abnormality, or folate deficiency. In practical terms, adolescents with normal BMI may still require hematologic and micronutrient assessment depending on the clinical context, dietary risk, symptoms, and local screening priorities. Conversely, adolescents with overweight or obesity may require structured cardiometabolic follow-up, especially repeat BP measurement, in addition to nutritional counseling.

Taken together, these findings support a school-linked screening model that combines anthropometric, BP, hematologic, and micronutrient assessments. However, the results indicate that screening programs should be designed with clear post-screening pathways. Adolescents with elevated/hypertensive-range BP require repeat office BP measurements and, when indicated, further diagnostic evaluation rather than immediate labeling as having hypertension (11). Similarly, adolescents with anemia, low ferritin concentration, vitamin B12 abnormality, or folate deficiency require clinical assessment to confirm the abnormality, evaluate the etiology, and guide treatment. In this context, school-linked screening should be viewed not as a standalone diagnostic intervention but as an entry point into structured adolescent preventive care.

Study Limitations

This study has several limitations. First, the study employed a retrospective cross-sectional design and was conducted within a single school-linked screening pathway, limiting causal inference and generalizability. Second, although 1,264 adolescents were reached and informed during school-based outreach, individual-level clinical and laboratory data were available for only 287 adolescents who completed the pediatric outpatient screening pathway. Therefore, the analytic cohort may not represent all students reached during outreach, and selection bias cannot be excluded. The analytic sample was also smaller than the sample size calculated for precise population-level prevalence estimation. For this reason, the reported values should be interpreted as frequencies within the analytic cohort of attendees rather than as population prevalence estimates.

Third, BP was measured during a single screening encounter using an oscillometric device. Although measurements were performed by the same clinician using an appropriately sized cuff after a period of seated rest, repeat measurements on separate visits and ambulatory BP monitoring were not available. Therefore, elevated or hypertensive-range BP values should be interpreted as screening findings rather than confirmed hypertension. Fourth, although adolescents with known chronic diseases were excluded, data on medication use, vitamin or iron supplementation, menstrual history, dietary intake, socioeconomic indicators, inflammatory markers, and follow-up outcomes were not systematically collected. These unmeasured factors may have influenced hematologic, micronutrient, and BP findings. Finally, ferritin was interpreted as an iron-store marker, but inflammatory status was not assessed; therefore, ferritin results should be interpreted in the context of screening rather than for definitive etiologic classification.

Despite these limitations, the study provides clinically relevant school-linked screening data from an adolescent cohort, evaluated using standardized anthropometric and BP definitions, along with hematologic and micronutrient measurements. The inclusion of confidence intervals, sex-

and BMI category-based comparisons, and multivariable analysis for elevated/hypertensive-range BP improves the transparency and interpretability of the findings. The results may inform more structured screening and referral pathways for adolescents that combine cardiometabolic and nutritional assessments.

Conclusion

Among adolescents who completed a school-linked clinical and laboratory screening pathway, excess weight, elevated or hypertensive-range BP, anemia, low ferritin concentration, vitamin B12 abnormalities, and folate deficiency were frequently detected. Girls had a higher burden of anemia and iron-related abnormalities, whereas boys had higher frequencies of abdominal obesity and abnormal BP screening findings. Higher BMI SDS and male sex were independently associated with elevated or hypertensive-range BP. These findings support integrating repeat BP assessment and targeted hematologic and micronutrient evaluation into school-linked adolescent health screening pathways, with particular attention to cardiometabolic follow-up in adolescents with higher BMI and to iron status evaluation in girls.

Ethics

Ethics Committee Approval: The study was approved by an University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital Institutional Non-Interventional Scientific Research Ethics Committee (approval no: 222-2026, date: 30.04.2026).

Informed Consent: The requirement for informed consent was waived because the study was retrospective and utilized anonymized data collected from routine screening.

Authorship Contributions

Concept: H.U.H., D.B., Design: H.U.H., D.B., Data Collection or Processing: H.U.H., E.O., O.H., M.B., Analysis or Interpretation: H.U.H., Literature Search: H.U.H., E.O., O.H., M.B., Writing: H.U.H., E.O., O.H., M.B., D.B.

Conflict of Interest: No conflicts of interest were declared by the authors.

Financial Disclosure: This study received no financial support.

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

References

1. Sawyer SM, Afifi RA, Bearinger LH, et al. Adolescence: a foundation for future health. Lancet. 2012;379:1630-40.
2. GBD 2021 Adolescent BMI Collaborators. Global, regional, and national prevalence of child and adolescent overweight and obesity, 1990-2021, with forecasts to 2050: a forecasting study for the Global Burden of Disease Study 2021. Lancet. 2025;405:785-812.
3. Ruan X, Zhu A, Wang T, et al. Global prevalence of hypertension in children and adolescents younger than 19 years: a systematic review and meta-analysis. JAMA Pediatr. 2025;179:987-99.
4. Glenning JP, Sheeran F, Quinlan C, Mynard JP. School-based screening for high blood pressure in children and adolescents: a systematic scoping review. Hypertens Res. 2026;49:1304-14.
5. Rietz H, Svärdhagen G, Wuehl E, Lurbe E, Brunström M. Prevalence of high blood pressure and hypertension among children and adolescents in Europe: a systematic review and meta-analysis. Arch Dis Child. 2026;111:264-71.
6. Culha Hosceylan I, Kallemoglu T, Basibuyuk M, Karabayir N. Iron and vitamin B12 levels in school-age children. J Pediatr Hematol Oncol. 2025.
7. Türkiye Cumhuriyeti Sağlık Bakanlığı, Dünya Sağlık Örgütü Avrupa Bölge Ofisi. Türkiye Çocukluk Çağı Obezite Araştırması COSI-TUR 2022. Ankara; 2022. Available from: https://hsgm.saglik.gov.tr/depo/birimler/saglikli-beslenme-ve-hareketli-hayat-db/Dokumanlar/Kitaplar/Turkiye_Cocukluk_Cagi_Obezite_Arastirmasi_2022.pdf. Accessed 3 June 2026.
8. Ay Akdağ S, Mercan Başpınar M, Basat O. The frequency of obesity, hypertension, and anemia in school-aged children: a retrospective cross-sectional study from a family health center. Turk J Fam Pract. 2021;25:69-75.
9. Barlow SE; Expert Committee. Expert committee recommendations regarding the prevention, assessment, and treatment of child and adolescent overweight and obesity: summary report. Pediatrics. 2007;120(Suppl4):S164-92.
10. Mramba L, Ngari M, Mwangome M, et al. A growth reference for mid upper arm circumference for age among school age children and adolescents, and validation for mortality: growth curve construction and longitudinal cohort study. BMJ. 2017;358:j3423.
11. Flynn JT, Kaelber DC, Baker-Smith CM, et al.; Subcommittee on Screening and Management of high Blood Pressure in Children. Clinical practice guideline for screening and management of high blood pressure in children and adolescents. Pediatrics. 2017;140:e20171904. Erratum in: Pediatrics. 2017;140:e20173035. Erratum in: Pediatrics. 2018;142:e20181739.
12. World Health Organization. Haemoglobin concentrations for the diagnosis of anaemia and assessment of severity. Geneva: World Health Organization; 2011. Available from: https://www.who.int/publications/i/item/WHO-NMH-NHD-MNM-11.1. Accessed June 3, 2026.

13. World Health Organization. WHO guideline on use of ferritin concentrations to assess iron status in individuals and populations. Geneva: World Health Organization; 2020. Available from: https://www.who.int/publications/i/item/9789240000124. Accessed June 3, 2026.
14. World Health Organization. Serum and red blood cell folate concentrations for assessing folate status in populations. Geneva: World Health Organization; 2015. Available from: https://www.who.int/publications/i/item/WHO-NMH-NHD-EPG-15.01. Accessed June 3, 2026.
15. Ferrari M, Mistura L, Patterson E, et al; HELENA Study Group. Evaluation of iron status in European adolescents through biochemical iron indicators: the HELENA study. Eur J Clin Nutr. 2011;65:340-9.
16. Weyand AC, Chaitoff A, Freed GL, Sholzberg M, Choi SW, McGann PT. Prevalence of iron deficiency and iron-deficiency anemia in US females aged 12-21 years, 2003-2020. JAMA. 2023;329:2191-3.
17. Sürücü Kara İ, Aydın Peker N, Dolğun İ, Mertoğlu C. Vitamin B12 level in children. J Curr Pediatr. 2023;21:127-34.
18. Kumar KJ, Balaji S, Kulkarni P, Manjunath VG, Nandish HR. Vitamin B12 status in rural adolescent school girls in mysuru, India: a community-based cross-sectional study. Indian J Community Med. 2023;48:794-7.
19. Kreusler P, Vogel M, Willenberg A, et al. Folate and cobalamin serum levels in healthy children and adolescents and their association with age, sex, BMI and socioeconomic status. Nutrients. 2021;13:546.
20. Vinker-Shuster M, Nakar-Weinstein A, Topf-Olivestone C, et al. Low utility of serum folic acid blood tests in healthy children and adolescents, a nationwide cohort. Eur J Pediatr. 2024;183:5173-9.

Original Article

DOI: 10.4274/haseki.galenos.2026.52714
Med Bull Haseki 2026;64(4):324-332

Ocular Surface and Meibomian Gland Phenotypes in Early-onset versus Late-onset Rheumatoid Arthritis: A Retrospective, Cross-sectional Comparative Imaging Study

Iffet Merve Ucar Baytaroglu[1], Ata Baytaroglu[2]

[1]*Usak Training and Research Hospital, Clinic of Rheumatology, Usak, Türkiye*
[2]*Usak Training and Research Hospital, Clinic of Ophthalmology, Usak, Türkiye*

Abstract

Aim: Ocular surface involvement in rheumatoid arthritis (RA) is well documented; however, the role of age at onset has not been characterized using modern topographic imaging. We compared ocular surface and meibomian gland phenotypes among early-onset RA (EORA), late-onset RA (LORA), and healthy controls, and related these phenotypes to systemic inflammatory and serologic features.

Methods: This single-center, retrospective, cross-sectional study (October 2022-September 2025) enrolled 126 age-matched participants: EORA, LORA, and healthy controls (n=42 per group). Rheumatoid arthritis was diagnosed according to the 2010 American College of Rheumatology/European League Against Rheumatism criteria. All participants completed the Ocular Surface Disease Index questionnaire and underwent slit-lamp biomicroscopy, Schirmer I testing, and Sirius+ topography to assess non-invasive break-up time, tear film thickness, and infrared meibography. 28-joint Disease Activity Score, rheumatoid factor, anti-cyclic citrullinated peptide, and antinuclear antibody status were recorded. Kruskal-Wallis test with Dunn-Bonferroni post-hoc comparisons, and chi-square or Fisher's exact tests were applied ($p<0.05$).

Results: Meibomian gland atrophy showed a trend toward higher values in EORA than in controls [33.0% (interquartile range 25.0-43.0) vs. 18.0% (10.0-35.8); $p=0.076$ after Bonferroni correction]. Ocular Surface Disease Index differed markedly across groups (EORA 41.0; LORA 15.0; controls 7.0; $p<0.001$). Tear film thickness was greatest in LORA (240.0 µm) and lowest in EORA (180.0 µm; $p=0.049$). Seropositivity was highest in EORA (80.0%) compared with LORA (63.2%) and controls (23.1%; $p<0.001$). Smoking was more frequent in LORA (28.6%) than in EORA (4.8%; $p=0.011$). 28-joint Disease Activity Score, erythrocyte sedimentation rate, and C-reactive protein did not differ between RA groups.

Conclusion: Despite comparable systemic disease activity, EORA and LORA appear to represent two distinct ocular surface phenotypes. Early-onset RA was associated with symptomatic presentation, seropositivity, and a trend toward greater meibomian gland atrophy, whereas LORA showed tear film instability, higher smoking exposure, and lower symptom reporting. Objective meibography, rather than symptom scoring alone, may help avoid underdiagnosis in older-onset RA.

Keywords: Arthritis, rheumatoid, age of onset, dry eye syndromes, meibomian glands, meibomian gland dysfunction, tear film

Corresponding Author: Ata Baytaroglu, MD, Usak Training and Research Hospital, Clinic of Ophthalmology, Usak, Türkiye
E-mail: baytaroglua@gmail.com **ORCID:** orcid.org/0000-0003-1993-0403
Received: 30.11.2025 **Accepted:** 27.07.2026 **Publication Date:** 24.09.2026

Cite this article as: Ucar Baytaroglu IM, Baytaroglu A. Ocular surface and meibomian gland phenotypes in early-onset versus late-onset rheumatoid arthritis: a retrospective, cross-sectional comparative imaging study. Med Bull Haseki. 2026;64(4):324-332

Introduction

Rheumatoid arthritis (RA) is a chronic, systemic inflammatory disease that primarily affects the synovial joints but also injures highly vascularized extra-articular tissues, including the ocular surface (1). Recent epidemiologic data place the global RA prevalence between 0.5% and 1.0% and confirm a steady rise in older-onset cases as populations age (2). Two clinical entities are now recognized on the basis of age at disease onset: [early-onset RA (EORA), onset <60 years] and [late-onset RA (LORA), onset ≥60 years], and recent reviews position LORA as a pathogenically distinct subset driven by immunosenescence, inflammaging, and clonal hematopoiesis rather than by classical autoimmunity (3,4). A 2026 systematic review and meta-analysis of more than 5,000 patients reported persistently higher post-treatment 28-joint Disease Activity Score (DAS28) scores and lower biologic-induced remission rates in LORA, supporting the view that age at onset modifies both immune phenotype and therapeutic response (5).

Ocular surface and meibomian gland involvement in RA has been described, but the relative contribution of EORA versus LORA remains poorly characterized in modern imaging studies. Comparative work to date has focused on joint and systemic features, with sonographic and clinical studies suggesting more active synovitis and a higher comorbidity burden in LORA, while autoantibody load and erosive disease tend to predominate in EORA (6,7). The eye-related literature has largely treated RA as a single disease, with ocular involvement reported in 28-55% of patients and ranging from mild keratoconjunctivitis sicca to vision-threatening peripheral ulcerative keratitis (8,9). Quantitative meibography, which directly visualizes acinar atrophy, has rarely been applied to the EORA versus LORA comparison, and the interaction with smoking exposure, an established environmental driver of citrullination and ocular surface damage, has not been systematically addressed (10).

We hypothesized that despite comparable systemic disease activity, EORA and LORA patients would display measurably different ocular surface phenotypes, with EORA showing more aggressive meibomian gland atrophy and higher subjective symptom burden and LORA showing tear film instability that is partially masked by reflex tearing related to environmental smoking. The aim of this retrospective, age-matched, single-center study was to compare ocular surface and meibomian gland imaging parameters between EORA, LORA, and healthy controls and to relate the findings to the autoantibody profile, smoking status, and DAS28-based disease activity. We aimed to provide phenotype-specific guidance for ophthalmologists managing RA patients, particularly the under-recognized LORA subgroup.

Materials and Methods

Compliance with Ethical Standards

This study was approved by the Usak University Non-Interventional Studies Ethics Committee (approval number: 907-907-12, date: 23.10.2025) and conducted in accordance with the principles of the Declaration of Helsinki. All patient identifiers were removed before analysis. The authors declare no conflict of interest. The study received no external funding.

Study Design

This was a single-center, retrospective, cross-sectional, comparative imaging study conducted in the Departments of Ophthalmology and Rheumatology at Usak University Training and Research Hospital between October 2022 and September 2025. Records of consecutive RA patients and age-matched non-RA controls were screened and examined for refractive evaluation. The detailed participant flow is shown in Figure 1.

Participants and Eligibility Criteria

Three groups were enrolled: EORA (disease onset before 60 years of age, n=42), LORA (disease onset at or after 60 years of age, n=42), and age-matched healthy controls (n=42). The diagnosis of RA was based on the 2010 American College of Rheumatology/European League Against Rheumatism classification criteria, and patients had at least one year of follow-up. Inclusion required age ≥18 years and the ability to complete the Ocular Surface Disease Index (OSDI) questionnaire and Sirius+ imaging. Exclusion criteria were spherical equivalent exceeding ±3.00 D, prior ocular or eyelid surgery, long-term contact-lens wear, active ocular infection or inflammation other than RA-related changes, allergic or vernal/atopic keratoconjunctivitis, use of topical ocular medications other than non-preserved artificial tears within the preceding 3 months, and any corneal pathology that would interfere with topographic imaging. Only the right eye of each participant was analyzed to avoid inter-eye correlation.

Ophthalmic Assessment

All participants underwent a standardized ophthalmic examination by a single experienced clinician (A.B.), which included best-corrected visual acuity, autorefraction, slit-lamp biomicroscopy of the lids, conjunctiva, cornea, and tear film, intraocular pressure measurement, fluorescein staining with non-preserved dye, and the OSDI questionnaire. Schirmer I testing was performed without anesthesia, and values of <5 mm/5 min were recorded as positive for aqueous tear deficiency. Non-invasive break-up time and tear film thickness were measured using the Sirius+ topography system (CSO, Florence, Italy) with Phoenix software (version 4.1.3.1).

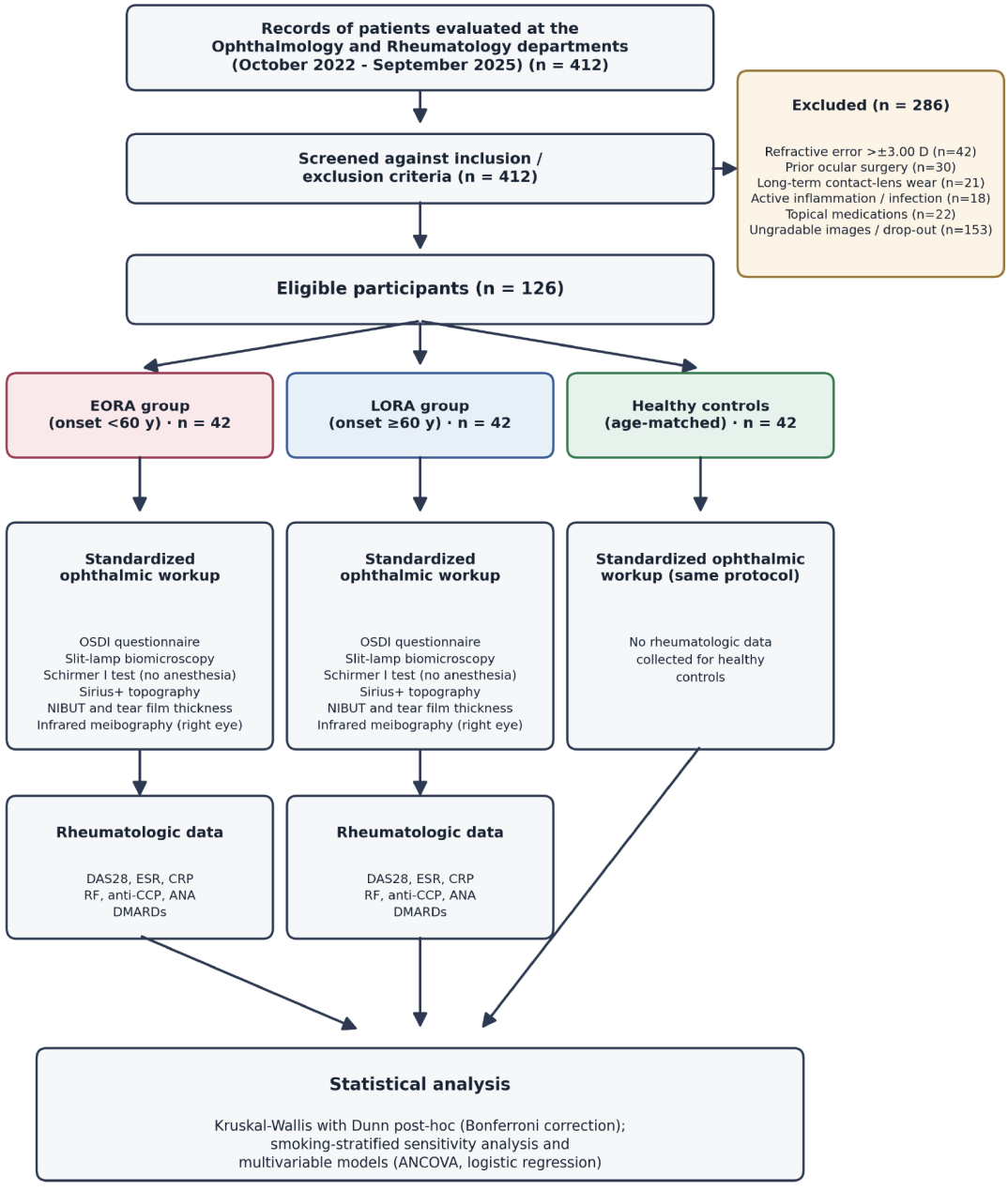

Figure 1. Study flow diagram. Flow of patients screened, excluded, enrolled, and analyzed in the EORA, LORA, and control groups. Standardized ophthalmic and rheumatologic workups were applied across the rheumatoid arthritis groups; healthy controls underwent the same ophthalmic protocol without rheumatologic data collection

EORA: Early-onset rheumatoid arthritis, LORA: Late-onset rheumatoid arthritis, OSDI: Ocular Surface Disease Index, NIBUT: Non-invasive break-up time, DAS28: 28-joint Disease Activity Score, ESR: Erythrocyte sedimentation rate, CRP: C-reactive protein, RF: Rheumatoid factor, anti-CCP: Anti-cyclic citrullinated peptide, ANA: Antinuclear antibody, DMARD: Disease-modifying anti-rheumatic drug

Meibography and Gland-atrophy Quantification

Infrared meibography was acquired using the meibography module embedded in the Sirius+ platform; the percentage of meibomian gland atrophy was automatically generated by the device's analysis software (Phoenix v.4.1.3.1) within the user-specified boundaries. The automatic atrophy estimation provided by the Sirius+ device shows good intra-device repeatability and segmentation performance comparable to recently validated convolutional-neural-network-based

meibography algorithms reported by Wang et al. (11) and Yesilirmak et al. (12), with semantic-segmentation Jaccard indices in the same range as those of expert graders (13). Each acquisition was screened for image quality (proper lid eversion, focus, and absence of motion blur) before automatic analysis, and unusable images were discarded and retaken.

Rheumatologic Assessment

For RA patients, disease duration, age at diagnosis, current and prior exposure to disease-modifying anti-rheumatic drugs (DMARDs), smoking status, and comorbidities were recorded. Laboratory parameters included rheumatoid factor (RF), anti-cyclic citrullinated peptide (anti-CCP), antinuclear antibody (ANA), erythrocyte sedimentation rate (ESR), and C-reactive protein (CRP) (Tables 1 and 2). Disease activity was assessed using the DAS28. Rheumatoid factor was measured by nephelometric assay and regarded as positive at the laboratory reference cut-off (>14 IU/mL); anti-CCP antibodies were measured by enzyme immunoassay and regarded as positive above the manufacturer's threshold (>17 U/mL). The same assays and cut-offs were applied to all participants, including controls, so that seropositivity in the control group reflects RF positivity at the standard diagnostic cut-off rather than low-titre or borderline reactivity. In the control group, serologic results were available only for participants for whom testing had been requested during routine evaluation, and this subset defined the denominator for control serology.

Statistical Analysis

An a priori power calculation for a one-way analysis of variance with three groups was conducted using G*Power v.3.1.9.2. An effect size of f=0.30, an alpha=0.05, and a power=0.80 indicated a minimum of 37 subjects per group; 42 subjects per group were enrolled to allow for exclusions. Distributional assumptions were assessed with Shapiro-Wilk testing; all primary continuous outcomes deviated from normality and were therefore analyzed with the Kruskal-Wallis test followed by Dunn post-hoc pairwise comparisons with Bonferroni correction. Categorical variables were compared using the chi-square or Fisher's exact test, as appropriate. Pre-specified sensitivity analyses were performed: a smoking-stratified Kruskal-Wallis comparison of tear film thickness, meibomian gland atrophy, and OSDI between EORA and LORA; and multivariable analysis of covariance and logistic regression adjusting for smoking and seropositivity (Tables 3 and 4). The analysis of covariance modeled each continuous ocular surface outcome (meibomian gland atrophy, tear film thickness, and OSDI), with group as the factor and smoking and seropositivity as covariates; logistic regression modeled seropositivity with group and smoking as predictors. Effect estimates [beta coefficients or odds ratios (ORs)] with 95% confidence intervals (CIs) are reported in Table 5. These multivariable models were secondary and were not corrected for multiple comparisons; thus, they complement rather than replace the Bonferroni-corrected primary analyses. Because of the moderate sample size and the marked imbalance in smoking prevalence between groups (4.8% in EORA vs. 28.6% in LORA), the multivariable models were considered supportive rather than demonstrating causality; results are reported with effect estimates and 95% CIs where appropriate. $p<0.05$ was considered statistically significant. Analyses were performed using GraphPad Prism v10.6.1 for macOS (GraphPad Software, Boston, MA, USA).

Results

A total of 126 participants were included (EORA n=42, LORA n=42, controls n=42). The median age was similar across groups [62.0 (60.5-63.5) years vs. 66.0 (61.0-71.0) years vs. 64.0 (62.0-66.0) years; p=0.06]. Demographic, serologic, and ocular surface variables are summarized in Tables 1-3 and Figure 2.

Table 1. Demographic, systemic, and ocular surface characteristics of early-onset (EORA), late-onset (LORA), and control groups

Variable	EORA (n=42)	LORA (n=42)	Controls (n=42)	p-value
Age (years)	62.0 (60.5-63.5)	66.0 (61.0-71.0)	64.0 (62.0-66.0)	0.060
ESR (mm/h)	21.0 (14.5-25.0)	21.0 (14.0-31.5)	n.a.	0.86†
CRP (mg/L)	6.2 (3.8-20.4)	6.5 (2.2-19.8)	n.a.	0.70†
DAS28	3.8 (2.7-4.6)	3.2 (2.7-4.1)	n.a.	0.59†
Meibomian gland atrophy (%)	33.0 (25.0-43.0)	26.0 (22.0-39.5)	18.0 (10.0-35.8)	0.035*
Tear film thickness (μm)	180.0 (115.0-230.0)	240.0 (200.0-315.0)	200.0 (180.0-245.0)	0.049*
NIBUT (s)	9.0 (6.0-12.0)	10.0 (6.0-12.0)	11.0 (7.3-14.0)	0.353
OSDI score	41.0 (18.5-45.5)	15.0 (11.0-24.5)	7.0 (3.3-12.0)	<0.001*

Values are median (interquartile range). Three-group comparisons used the Kruskal-Wallis test; †Indicates a two-group Mann-Whitney U test (EORA vs. LORA only) for biomarkers limited to the rheumatoid arthritis groups. *p<0.05
EORA: Early-onset rheumatoid arthritis, LORA: Late-onset rheumatoid arthritis, ESR: Erythrocyte sedimentation rate, CRP: C-reactive protein, DAS28: 28-joint Disease Activity Score, NIBUT: Non-invasive break-up time, OSDI: Ocular Surface Disease Index

Table 2. Distribution of categorical variables across groups				
Variable	**EORA (%)**	**LORA (%)**	**Controls (%)**	**p-value**
Male sex	23.8	26.2	38.1	0.872
Smoking	4.8	28.6	21.4	0.011*
Any comorbid disease	26.2	28.6	31.0	0.321
Diabetes mellitus	7.1	11.9	7.1	0.441
Hypertension	16.7	19.0	19.0	0.553
Coronary artery disease	4.8	7.1	23.8	0.107
Seropositivity (RF and/or anti-CCP)	80.0	63.2	23.1	<0.001*
RF positivity	73.3	63.2	23.1	<0.001*
Anti-CCP positivity	60.0	52.6	0.0	<0.001*
ANA positivity	40.0	31.6	15.4	0.193
Extra-articular involvement	7.1	0.0	0.0	0.076
Prednisolone use	21.4	16.7	0.0	0.300
Methotrexate use	14.3	23.8	0.0	0.510
Leflunomide use	19.0	14.3	0.0	0.296
Biologic DMARD use	4.8	2.4	0.0	0.571
Hydroxychloroquine use	4.8	9.5	0.0	0.672
Sulfasalazine use	4.8	0.0	0.0	0.187
Limbal hyperemia (score ≥1)	16.7	19.0	26.2	0.140
Positive Schirmer test (<5 mm/5 min)	9.5	7.1	0.0	0.030*

Categorical variables compared with the chi-square test or Fisher exact test as appropriate. *p<0.05
Serologic percentages were calculated among participants with available serology; in the control group, this denominator comprised only those tested during routine evaluation, and control seropositivity denotes rheumatoid factor positivity at the standard laboratory cut-off (>14 IU/mL), not low-titre reactivity
EORA: Early-onset rheumatoid arthritis, LORA: Late-onset rheumatoid arthritis, RF: Rheumatoid factor, anti-CCP: Anti-cyclic citrullinated peptide, ANA: Antinuclear antibody, DMARD: Disease-modifying anti-rheumatic drug

Table 3. Dunn post-hoc pairwise comparisons (Bonferroni correction) for variables with significant Kruskal-Wallis results				
Variable	**Comparison**	**Z**	**p-value**	**Bonferroni-adjusted p**
Meibomian gland atrophy (%)	EORA vs. LORA	0.56	0.578	1.000
	EORA vs. controls	2.24	0.025	0.076
	LORA vs. controls	1.99	0.047	0.140
Tear film thickness (μm)	EORA vs. LORA	2.24	0.025	0.075
	EORA vs. controls	1.55	0.121	0.363
	LORA vs. controls	1.40	0.163	0.490
OSDI score	EORA vs. LORA	2.37	0.018	0.055
	EORA vs. controls	>3.6	<0.001	<0.001*
	LORA vs. controls	3.40	<0.001	0.002*

Only variables with a significant Kruskal-Wallis p<0.05 entered the post-hoc analysis. *p<0.05 after Bonferroni correction
EORA: Early-onset rheumatoid arthritis, LORA: Late-onset rheumatoid arthritis, OSDI: Ocular Surface Disease Index

Table 4. Smoking-stratified sensitivity analysis of ocular surface outcomes between EORA and LORA non-smokers			
Variable	**EORA non-smokers (n=40)**	**LORA non-smokers (n=30)**	**p-value**
Meibomian gland atrophy (%)	33.5 (26.0-44.0)	25.0 (21.0-38.0)	0.046*
Tear film thickness (μm)	182.0 (118.0-232.0)	232.0 (198.0-306.0)	0.062
NIBUT (s)	9.0 (6.0-12.0)	10.5 (7.0-13.0)	0.412
OSDI score	41.5 (19.0-46.0)	16.0 (11.5-25.0)	<0.001*

Pre-specified sensitivity analysis restricted to never- or ex-smokers. Two-group comparison performed with the Mann-Whitney U test. *p<0.05
EORA: Early-onset rheumatoid arthritis, LORA: Late-onset rheumatoid arthritis, NIBUT: Non-invasive break-up time, OSDI: Ocular Surface Disease Index

Table 5. Multivariable models of ocular surface outcomes and seropositivity, adjusted for smoking and seropositivity				
Model	**Outcome/predictor**	**Estimate**	**Adjusted estimate (95% CI)**	**p-value**
ANCOVA	OSDI score (EORA vs. LORA)	β	27.7 (22.4 to 33.0)	<0.001
ANCOVA	Tear film thickness, μm (EORA vs. LORA)	β	-57.9 (-78.5 to -37.2)	<0.001
ANCOVA	Meibomian gland atrophy, % (EORA vs. LORA)	β	4.3 (0.1 to 8.5)	0.045
Logistic regression	Seropositivity: EORA vs. others	OR	5.7 (2.3 to 14.1)	<0.001
Logistic regression	Seropositivity: smoking	OR	1.3 (0.5 to 3.4)	0.58

Analysis of covariance (ANCOVA) estimates are beta coefficients (β) for the EORA versus LORA contrast (reference: LORA), adjusted for smoking and seropositivity; positive values indicate higher outcomes in EORA and the tear film thickness coefficient is negative because values were lower in EORA. Logistic regression estimates are odds ratios for seropositivity (rheumatoid factor and/or anti-CCP), adjusted for smoking. These secondary models were not corrected for multiple comparisons and complement the Bonferroni-corrected analyses in Table 3. CI: Confidence interval, OSDI: Ocular Surface Disease Index, EORA: Early-onset rheumatoid arthritis, LORA: Late-onset rheumatoid arthritis

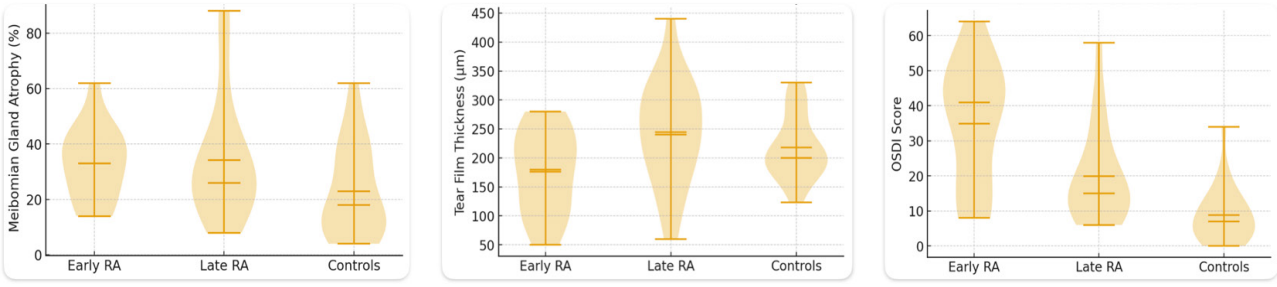

Figure 2. Violin plots of ocular surface parameters across the three groups. Distribution of meibomian gland atrophy (%), tear film thickness (μm), non-invasive break-up time (s), and Ocular Surface Disease Index score across EORA, LORA, and healthy controls. Boxes within violins represent medians and IQRs; whiskers show 1.5 × IQR. Three-group comparisons were performed with the Kruskal-Wallis test

EORA: Early-onset rheumatoid arthritis, LORA: Late-onset rheumatoid arthritis, IQR: Interquartile range, CI: Confidence interval, RA: Rheumatoid arthritis

Systemic disease activity did not differ between EORA and LORA [DAS28: 3.8 (2.7-4.6) vs. 3.2 (2.7-4.1), p=0.59; ESR, p=0.86; CRP, p=0.70]. Seropositivity (RF and/or anti-CCP) differed across groups (Figure 3): it was highest in EORA (80.0%), intermediate in LORA (63.2%), and lowest in controls (23.1%; p<0.001). Rheumatoid factor and anti-CCP each demonstrated the same gradient (both p<0.001). Antinuclear antibody positivity was highest in EORA (40.0%) but did not reach statistical significance (p=0.193).

Meibomian gland atrophy tended to be higher in the EORA group [33.0% (25.0-43.0)] than in controls [18.0% (10.0-35.8)]; overall Kruskal-Wallis p=0.035. The pairwise comparison did not remain significant after Bonferroni correction (EORA vs. control p=0.076). Tear film thickness differed across groups (p=0.049), with the highest median value in LORA [240.0 μm (200.0-315.0)] and the lowest in EORA [180.0 μm (115.0-230.0)]; no pairwise comparisons remained significant after Bonferroni correction (Table 3). Non-invasive break-up time did not differ between groups (p=0.353).

Ocular Surface Disease Index showed the largest between-group separation (p<0.001). EORA patients reported the highest symptom burden [median 41.0 (18.5-45.5)], LORA patients reported intermediate scores [median 15.0 (11.0-24.5)], and controls reported minimal symptoms [median 7.0 (3.3-12.0)]. Both RA groups had significantly higher OSDI scores than controls (adjusted p<0.001 for both). The EORA versus LORA comparison was borderline significant (adjusted p=0.055).

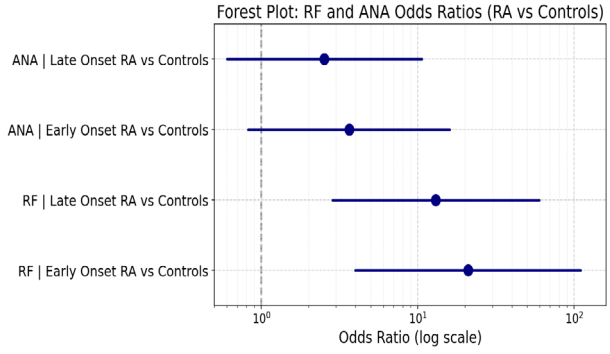

Figure 3. Forest plot of odds ratios for serologic positivity [rheumatoid factor (RF) and antinuclear antibody (ANA)] across rheumatoid arthritis subtypes versus healthy controls Odds ratios for RF and ANA positivity in EORA and LORA versus healthy controls, plotted on a logarithmic scale with 95% confidence intervals. RF positivity discriminates both RA subtypes from controls. ANA positivity shows a non-significant trend toward higher rates in RA

EORA: Early-onset rheumatoid arthritis, LORA: Late-onset rheumatoid arthritis, RA: Rheumatoid arthritis

Smoking prevalence differed across groups (4.8% in EORA, 28.6% in LORA, and 21.4% in controls; p=0.011). A positive Schirmer test was observed only in the RA groups (9.5% in EORA, 7.1% in LORA, and 0% in controls; p=0.030). In the pre-specified, smoking-stratified analysis restricted to non-smokers, EORA still exhibited higher meibomian gland atrophy and OSDI scores than LORA (Table 4), suggesting that the EORA phenotype was not explained by lower smoking exposure. In the multivariable models (Table 5), the estimates were directionally consistent with the unadjusted comparisons. After adjustment for smoking and seropositivity, OSDI remained higher in EORA than in LORA (adjusted β 27.7, 95% CI 22.4 to 33.0; p<0.001), and tear film thickness remained lower in EORA (adjusted β -57.9, 95% CI -78.5 to -37.2; p<0.001). The difference in meibomian gland atrophy was modest, with a CI that approached the null (adjusted β 4.3, 95% CI 0.1 to 8.5; p=0.045), consistent with a trend rather than a reliable independent effect. In the logistic model, seropositivity was associated with EORA (adjusted OR 5.7, 95% CI 2.3 to 14.1; p<0.001) but not with smoking (OR 1.3, 95% CI 0.5 to 3.4; p=0.58). Other comorbidities, DMARD exposure, and ANA status did not differ between EORA and LORA (Figure 3).

Discussion

An age-matched comparison of ocular surface and meibomian gland phenotypes between EORA and LORA cohorts using objective topographic imaging has not hitherto been reported. The pattern that emerged from our data was unforeseen in two respects. Early-onset RA showed a trend toward greater meibomian gland atrophy and significantly higher OSDI scores than both LORA and controls, although DAS28, ESR, and CRP were comparable across RA subgroups; this difference in atrophy did not remain significant after correction for multiple comparisons and was therefore interpreted as a trend. LORA, despite a sixfold higher smoking prevalence, had the thickest tear film and the lowest symptom burden among RA subgroups. Seropositivity for RF or anti-CCP, while elevated in both groups, tracked the EORA gland-atrophy phenotype more closely than the LORA phenotype.

Our observation that systemic disease activity scores fail to predict ocular surface severity aligns with the meta-analysis by Yang et al. (5), which found that LORA achieved remission less often than younger-onset RA on biologic and targeted synthetic DMARDs at comparable baseline DAS28. It also fits the framework of Ma et al. (3), who place the EORA/LORA distinction at the level of immune mechanism rather than at the level of joint count or acute-phase reactants. From this point of view, the ocular surface may offer a second readout of disease biology: in EORA, reflecting a phenotype associated with long-standing, seropositive, B-cell-related autoimmunity; in LORA, reflecting an older immune background in which environmental exposures may play a comparatively larger role. These interpretations describe associations rather than demonstrated mechanisms, because the retrospective design cannot establish causality.

The high seropositivity rates in our EORA group (RF 73.3%, anti-CCP 60.0%) are consistent with, but do not establish, a B-cell-related contribution to meibomian gland injury. Anti-citrullinated protein antibodies are increasingly recognized as drivers of tissue damage beyond the synovium, and citrullinated targets have been identified in the lacrimal gland and conjunctival epithelium. The preferential involvement of meibomian glands, which are sebaceous and androgen-regulated, may reflect several mechanisms acting together: direct autoantibody binding (14,15), long-term cytokine exposure [tumor necrosis factor alpha, interleukin (IL)-6, IL-17], and loss of androgen-mediated trophic support. All three are well documented in RA (16,17). These pathways were not directly assessed in the present study. Long-standing seropositive RA has been associated with extra-articular manifestations and exocrine gland involvement (18), and Targońska-Stępniak et al.'s (19) comparison of LORA and EORA reported a similar immune-aggression gradient at the systemic level, although ocular outcomes were not measured.

The LORA presentation, in which mild symptoms persist despite the absence of dry eye findings, warrants closer examination. Cigarette smoke is a well-characterized ocular surface toxin that destabilizes the lipid layer, depletes goblet cells, and reduces meibomian gland density (20). It is conspicuous that the magnitude of these effects in chronic smokers does not appear to depend on smoking duration or nicotine dependence in a dose-dependent manner (20). If smoking exerted a uniform pressure to increase glandular atrophy in our LORA group, the higher atrophy observed in EORA would be consistent with an autoimmune-related signal; however, this requires validation in a prospective study. Smoking-related compensatory reflex tearing in LORA may also explain the greater tear film thickness despite poorer underlying tear stability, an interpretation supported by recent immunological data linking smoking to preferential induction of mucosal IgA anti-citrullinated protein antibodies (21). Tear film hyperosmolarity has previously been reported as a feature of RA-related dry eye, with disease activity, treatment, and age contributing independently to ocular surface signs (22,23). In our pre-specified smoking-stratified sub-analysis, the difference in OSDI between EORA and LORA persisted among non-smokers, with a similar but weaker trend for meibomian gland atrophy.

The sign-symptom dissociation we observed has direct clinical implications. Early-onset RA patients reported severe symptoms (median OSDI score of 41), whereas LORA patients reported only mild symptoms (OSDI score of 15) despite the presence of moderate gland atrophy. Two mechanisms could plausibly explain this. Early-onset RA patients may develop corneal neuropathic pain and central sensitization against a background of long-standing systemic inflammation, which would align with the increased corneal Langerhans-cell density and reduced corneal nerve fiber length reported by Bitirgen et al. (24) in RA. Conversely, LORA patients may have blunted corneal nerve sensitivity due to immunosenescence, resulting in underreporting of objective gland loss on questionnaires. Symptom scoring alone may, therefore, underestimate ocular surface disease in LORA, and direct meibography is a more reliable screening tool than symptom scoring. Deep-learning meibography pipelines now achieve expert-level segmentation accuracy (11-13) and should reduce operator dependence in routine use.

Our results also align with the gerontological view of difficult-to-treat RA from Lehoczki et al. (4), in which cumulative organ vulnerability, polypharmacy, and frailty determine outcomes more than acute-phase activity. From that perspective, the LORA ocular phenotype, with its lower autoantibody load and higher environmental exposure, is best managed with tobacco-cessation counseling and proactive lid-margin care, whereas EORA requires tighter coordination with rheumatology regarding biologic therapy and screening for secondary Sjögren overlap (16).

Our 40% ANA positivity rate in EORA, although not statistically significant, is clinically noteworthy. Combined with the higher frequency of positive Schirmer tests in EORA (9.5%) and the more pronounced subjective symptoms in EORA, this pattern suggests an under-recognized overlap with secondary Sjögren disease. Sullivan et al. (16) demonstrated that meibomian gland dysfunction is a near-universal finding in primary and secondary Sjögren syndrome. In our setting, the addition of meibography to standard ocular surface evaluation is therefore likely to identify patients who would benefit from referral for Sjögren work-up, even in the absence of overt sicca complaints. Our findings add to existing evidence on dry eye severity in autoimmune rheumatic disease (25,26) and on the role of sustained inflammation in lacrimal gland injury (27). Two recent practical reviews on LORA management offer additional context for tailoring DMARD therapy in this older subgroup, where ocular adverse effects and comorbidity load demand careful coordination with rheumatology (28,29).

Study Limitations

Several limitations should be acknowledged. The study design precludes causal inference and cannot capture the progression of meibomian gland atrophy over time; recruitment from a single center limits external validity, although demographic comparability between EORA and LORA mitigates concerns about internal validity. The marked smoking imbalance between RA subgroups, combined with a moderate sample size, prevented a fully powered multivariable adjustment; the pre-specified stratified and secondary regression analyses are therefore reported as supportive rather than confirmatory, and residual confounding by methotrexate or biologic therapy cannot be excluded, although DMARD distribution did not differ between RA subgroups (18). On the instrumentation side, the Sirius+ automatic gland-atrophy segmentation has been validated in healthy subjects but not specifically in autoimmune cohorts; recent deep-learning meibography studies do, however, suggest close agreement between automatic and manual grading (11-13). Tear cytokine profiling and corneal confocal microscopy were not available, and their inclusion would have strengthened the mechanistic discussion. Given these limitations, the study is, as far as we are aware, the first direct comparison of ocular surface and meibomian gland imaging between age-matched EORA, LORA, and control groups; the rheumatologic and ophthalmologic workups were standardized, the analysis was driven by an a priori powered design with pre-specified sensitivity analyses, and meibography acquisition followed validated protocols (30).

Conclusion

Our findings suggest that EORA and LORA may represent two distinct ocular surface phenotypes. EORA was characterized by symptomatic presentation and seropositivity, with a trend toward greater meibomian gland atrophy, whereas LORA was associated with tear film instability and comparatively few symptoms. Ocular surface screening in older patients with RA should, therefore, be guided by objective meibography rather than symptom scoring alone.

Ethics

Ethics Committee Approval: This study was approved by the Usak University Non-Interventional Studies Ethics Committee (approval number: 907-907-12, date: 23.10.2025).

Informed Consent: Written informed consent was waived because of the retrospective design.

Footnotes

Authorship Contributions

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

Conflict of Interest: The authors declared that there were no conflicts of interest.

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

References

1. Lee DM, Weinblatt ME. Rheumatoid arthritis. Lancet. 2001;358:903-11.
2. Uke P, Maharaj A, Adebajo A. A review on the epidemiology of rheumatoid arthritis: an update and trends from current literature. Best Pract Res Clin Rheumatol. 2025;39:102036.
3. Ma S, Wu X, Zhang X, Li K. Elderly-onset rheumatoid arthritis as a distinct entity: mechanisms, manifestations, and treatment. Curr Opin Immunol. 2026;99:102721.
4. Lehoczki A, Ungvari Z, Szappanos Á, Fekete M, Gunkl-Tóth L, Nagy G. Geroscience insights into difficult-to-treat rheumatoid arthritis: the role of unhealthy aging, comorbidity, and therapeutic complexity. Geroscience. 2026;48:2153-77.
5. Yang M, Xu F, Zhang B, Pi H. Age-stratified treatment response in rheumatoid arthritis: a systematic review, meta-analysis, and integrated genetic and single-cell evidence implicating IL6R/TYK2 signaling. Clin Rheumatol. 2026;45:2583-600.
6. Aziz NN, Hassan RM, Ibrahim RA, Afifi N, Thabet RN. Clinical and sonographic pattern of late-onset and early-onset rheumatoid arthritis: comparative study. Clin Med Insights Arthritis Musculoskelet Disord. 2026;19:11795441261429109.
7. Zimba O, Baimukhamedov C, Kocyigit BF. Late-onset rheumatoid arthritis: clinical features, diagnostic challenges, and treatment approaches. Rheumatol Int. 2025;45:152.
8. Ruiz-Lozano RE, Prida-Espaillat LA, Moll-Auais RA, et al. Ocular inflammatory manifestations in patients with rheumatoid arthritis. Reumatol Clin (Engl Ed). 2025;21:501891.
9. Turk MA, Hayworth JL, Nevskaya T, Pope JE. Ocular manifestations in rheumatoid arthritis, connective tissue disease, and vasculitis: a systematic review and metaanalysis. J Rheumatol. 2021;48:25-34.
10. Sparks JA, Karlson EW. The roles of cigarette smoking and the lung in the transitions between phases of preclinical rheumatoid arthritis. Curr Rheumatol Rep. 2016;18:15.
11. Wang J, Graham AD, Yu SX, Lin MC. Predicting demographics from meibography using deep learning. Sci Rep. 2022;12:15701.
12. Yesilirmak N, Okbay V, Yesilirmak Y, et al. Evaluation of meibomian gland dysfunction with deep learning model considering different datasets and gland morphology. Comput Biol Med. 2025;195:110679.
13. Swiderska K, Blackie CA, Maldonado-Codina C, Morgan PB, Read ML, Fergie M. A deep learning approach for meibomian gland appearance evaluation. Ophthalmol Sci. 2023;3:100334.
14. Vignesh AP, Srinivasan R. Ocular manifestations of rheumatoid arthritis and their correlation with anti-cyclic citrullinated peptide antibodies. Clin Ophthalmol. 2015;9:393-7.
15. Korkmaz C, Us T, Kaşifoğlu T, Akgün Y. Anti-cyclic citrullinated peptide (CCP) antibodies in patients with long-standing rheumatoid arthritis and their relationship with extra-articular manifestations. Clin Biochem. 2006;39:961-5.
16. Sullivan DA, Dana R, Sullivan RM, et al. Meibomian gland dysfunction in primary and secondary Sjögren syndrome. Ophthalmic Res. 2018;59:193-205.
17. Khandelwal P, Liu S, Sullivan DA. Androgen regulation of gene expression in human meibomian gland and conjunctival epithelial cells. Mol Vis. 2012;18:1055-67.
18. Sevimli N, Karadag R, Madenci E, et al. Ocular surface findings in patients with rheumatoid arthritis under methotrexate or biological agent therapy. Cutan Ocul Toxicol. 2017;36:48-51.
19. Targońska-Stępniak B, Grzechnik K, Kolarz K, Gągoł D, Majdan M. Systemic inflammatory parameters in patients with elderly-onset rheumatoid arthritis (EORA) and young-onset rheumatoid arthritis (YORA)-an observational study. J Clin Med. 2021;10:1204.
20. Ağın A, Kocabeyoğlu S, Çolak D, İrkeç M. Ocular surface, meibomian gland alterations, and in vivo confocal microscopy characteristics of corneas in chronic cigarette smokers. Graefes Arch Clin Exp Ophthalmol. 2020;258:835-41.
21. van Wesemael TJ, Morton FR, Heutz JW, et al. Smoking as a risk factor for rheumatoid arthritis: predominant association with IgA autoantibodies - comprehensive analysis of anti-modified protein antibodies with smoking and genetic risk factors in rheumatoid arthritis. Arthritis Res Ther. 2025;27:101.
22. Schargus M, Wolf F, Tony HP, Meyer-Ter-Vehn T, Geerling G. Correlation between tear film osmolarity, dry eye disease, and rheumatoid arthritis. Cornea. 2014;33:1257-61.
23. Ma W, Wang G, Li X, et al. Study of factors influencing dry eye in rheumatoid arthritis. J Ophthalmol. 2020;2020:5768679.
24. Bitirgen G, Kucuk A, Ergun MC, Satirtav G, Malik RA. Corneal nerve loss and increased Langerhans cells are associated with disease severity in patients with rheumatoid arthritis. Eye (Lond). 2023;37:2950-5.
25. Liu Y, Wu M, Ren Y, et al. Evaluation of dry eye severity and ocular surface inflammation in patients with autoimmune rheumatic diseases. Ocul Immunol Inflamm. 2024;32:2018-30.
26. Yu K, Bunya V, Maguire M, Asbell P, Ying GS; Dry Eye Assessment and Management Study Research Group. Systemic conditions associated with severity of dry eye signs and symptoms in the dry eye assessment and management study. Ophthalmology. 2021;128:1384-92.
27. Villani E, Galimberti D, Del Papa N, Nucci P, Ratiglia R. Inflammation in dry eye associated with rheumatoid arthritis: cytokine and in vivo confocal microscopy study. Innate Immun. 2013;19:420-7.
28. Yoshimi R, Nakajima H. The characteristics and management considerations of late-onset rheumatoid arthritis. Intern Med. 2024;63:3267-9.
29. Kojima M, Sugihara T, Kawahito Y, et al. Consensus statement on the management of late-onset rheumatoid arthritis. Mod Rheumatol. 2024;34:1095-102.
30. Finis D, Ackermann P, Pischel N, et al. Evaluation of meibomian gland dysfunction and local distribution of meibomian gland atrophy by non-contact infrared meibography. Curr Eye Res. 2015;40:982-9.

Original Article

DOI: 10.4274/haseki.galenos.2026.15238
Med Bull Haseki 2026;64(4):333-344

Semantic Architecture and Knowledge Evolution of Endoscopic Submucosal Dissection Research: A Computational Bibliometric Analysis

Huseyin Demir[1], Tuncay Akar[2], Suleyman Mertoglu[3], Merve Tekinarslan[4]

[1]*Izmir Katip Celebi University Faculty of Economics and Administrative Sciences, Department of Health Management, Izmir, Türkiye*
[2]*Mugla Sıtkı Kocman University Institute of Health Sciences, Department of Health Management, Mugla, Türkiye*
[3]*Izmir Provincial Directorate of Health, Izmir, Türkiye*
[4]*Gumushane University Faculty of Health Sciences, Department of Health Management, Gumushane, Türkiye*

Abstract

Aim: The overarching intellectual structure, dominant thematic clusters, and longitudinal evolution of the endoscopic submucosal dissection (ESD) literature have yet to be systematically mapped. This study aims to characterize the semantic and thematic architecture of ESD research through computational text analytics and to reveal latent structural patterns and temporal dynamics within this domain.

Methods: A corpus of 1,000 peer-reviewed publications was processed through an eight-stage computational pipeline that incorporated word embedding-based semantic analysis, unsupervised clustering, Latent Dirichlet Allocation (LDA) topic modeling, inter-topic network analysis, and temporal trend modeling. The primary outcome was characterization of the semantic architecture of ESD research, whereas secondary outcomes included topic relationships, temporal evolution, and model performance metrics.

Results: Semantic analysis revealed two dominant clusters: a technically oriented domain encompassing device innovation and procedural optimization and a clinically oriented domain addressing therapeutic outcomes, organ-specific protocols, and adverse event management. Topic modeling identified nine coherent thematic categories, with limited integration between technical and clinical research trajectories. Temporal analysis revealed a pronounced post-2024 shift toward upper-gastrointestinal applications, particularly esophageal and gastric ESD, with relative stagnation in colorectal investigations. The optimal semantic clustering solution achieved a silhouette score of 0.215, while the nine-topic LDA model demonstrated a coherence score of 0.39.

Conclusion: This study provides the first data-driven macro-structural characterization of the ESD literature, revealing a persistent bifurcation between technical innovation and clinical application streams. These findings offer a strategic framework to help clinicians, investigators, and curriculum developers identify priority areas for advancement in therapeutic endoscopy.

Keywords: Endoscopic submucosal dissection, treatment outcome, bibliometrics, semantics, cluster analysis, policy

Introduction

Endoscopic submucosal dissection (ESD), first introduced in Japan during the late 1990s (1), represents a major advancement in therapeutic endoscopy. Developed to overcome the limitations of endoscopic mucosal resection (EMR) (2,3), ESD enables en bloc resection of large and morphologically heterogeneous neoplastic lesions without lymphovascular invasion (4,5), conferring significant histopathological advantages in margin assessment and invasion depth determination. The inability of conventional EMR to achieve en bloc resection in lesions larger than 20 mm, combined with elevated recurrence

Corresponding Author: Huseyin Demir, MD, Izmir Katip Celebi University Faculty of Economics and Administrative Sciences, Department of Health Management, Izmir, Türkiye
E-mail: huseyin.demir@ikc.edu.tr **ORCID:** orcid.org/0000-0002-8990-7228
Received: 21.01.2026 **Accepted:** 17.07.2026 **Publication Date:** 24.09.2026

Cite this article as: Demir H, Akar T, Mertoglu S, Tekinarslan M. Semantic architecture and knowledge evolution of endoscopic submucosal dissection research: a computational bibliometric analysis. Med Bull Haseki. 2026;64(4):333-344

rates following piecemeal resection (6,7), underscores ESD's clinical value. Recent studies indicate that ESD is the best way to treat superficial gastric neoplasias and early gastric cancer (2). It is also being used more and more to treat esophageal squamous cell carcinoma and some colorectal cases, especially lesions that are larger than 20 mm or that are not suitable candidates for conventional polypectomy (8-10). Despite its clinical impact, ESD remains technically demanding, with complication rates of 3.5%, 3.3%, and 4.6% for gastric, esophageal, and colorectal applications, respectively (11-13). Submucosal fibrosis and limited operator expertise in Western settings have further constrained broader implementation, though structured training initiatives may mitigate these barriers (14,15).

The ESD literature spans diverse domains, including clinical efficacy, complication management, technological innovation, and organ-specific protocols (16,17). However, no study has examined this corpus at the macrostructural level to characterize thematic clustering, latent semantic architecture, or temporal evolution—the gap that limits the systematic understanding of the field's research organization and strategic priorities. Natural language processing (NLP) and unsupervised machine learning methodologies, particularly word embedding-based semantic analysis and Latent Dirichlet Allocation (LDA), offer robust frameworks for revealing latent thematic structures within large medical corpora (18). The thematic trajectory of ESD literature may further be interpreted through Rogers' diffusion of innovations framework (19), wherein early feasibility and instrumentation studies reflect initial adoption stages, while contemporary focus on outcomes, complication protocols, and organ-specific applications signals progression toward implementation and confirmation phases. We hypothesized that computational analysis of the ESD literature would reveal a structurally fragmented research landscape, organized along two parallel but largely disconnected trajectories: one centered on technical innovation and procedural optimization, and the other focused on clinical outcomes and organ-specific applications, with a discernible temporal shift toward upper gastrointestinal domains over the 2015-2025 period.

Unlike conventional bibliometric studies that primarily rely on citation counts, co-authorship networks, or keyword co-occurrence analyses, the present study integrates NLP with word embedding–based semantic analysis and LDA topic modeling to uncover the latent conceptual organization of the ESD literature. This integrated computational framework enables a deeper exploration of semantic relationships, thematic structures, and temporal knowledge evolution beyond the capabilities of traditional bibliometric approaches. In this context, the primary aim of this study is to systematically characterize the semantic, thematic, and temporal architecture of ESD research through computational text analytics, thereby revealing latent structural patterns and dynamics within this specialized domain. A secondary aim is to identify dominant thematic clusters, inter-topic relationships, and longitudinal research trends within the field. If these objectives are achieved, this analysis is expected to provide clinicians, investigators, and medical educators with an empirically grounded roadmap for prioritizing future research agendas, optimizing procedural training curricula, and informing health policy decisions related to ESD implementation.

Materials and Methods

Compliance with Ethical Standards

This study is a retrospective analysis of publicly available publication data retrieved from bibliometric databases. The analysis did not involve any human subjects, primary data collection, or interventions. Since all data used in this research were already in the public domain and freely accessible, and no identifiable personal information or sensitive data were processed, ethical approval from an institutional review board was not required.

Study Design

This study employed a retrospective computational bibliometric design based on scientific publications, integrating bibliometric methods with NLP and computational text analytics, including latent semantic analysis and topic modeling, to systematically map the intellectual structure and temporal evolution of the ESD literature. To achieve this objective, the analysis was conducted using an eight-step computational pipeline (Figure 1), enabling a systematic and multidimensional evaluation of the ESD research landscape.

The pipeline shown in Figure 1 consists of data collection, preprocessing, word embedding-based semantic vector generation, semantic clustering, topic modeling, inter-topic correlation analysis, temporal dynamics evaluation, model performance measurement, and visualization of results.

Step 1: Data Source, Search Strategy, and Research Unit

The data used in this study were obtained from the Web of Science (WoS) Core Collection (20), a widely recognized and trusted source for examining scientific literature in bibliometric and text analytics research. The study was conducted under the assumption that publications retrieved from the WoS database adequately reflect developments in the relevant scientific field. The unit of analysis for the study consists of the 1,000 most-cited articles identified

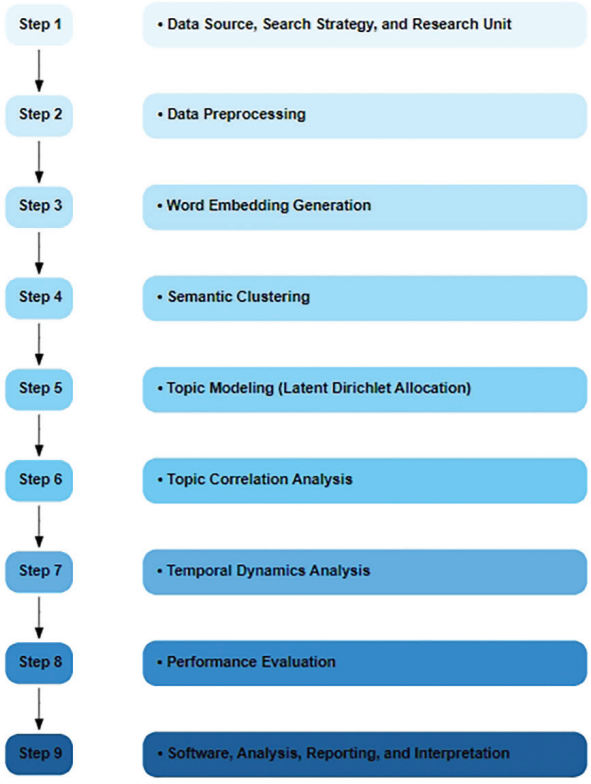

Figure 1. Eight-step computational workflow used for semantic and thematic analysis of the ESD literature. The workflow includes data collection, text preprocessing, Word2Vec-based document embedding, semantic clustering, LDA topic modeling, topic correlation analysis, temporal trend analysis, model performance evaluation, and data visualization
ESD: Endoscopic submucosal dissection, LDA: Latent Dirichlet Allocation

in the database and published in the last decade (2015-2025). The selection of the 1,000 most-cited publications within the 2015-2025 period was guided by both methodological and substantive considerations. Citation frequency serves as a widely validated proxy for scientific influence; highly cited publications disproportionately shape the conceptual vocabulary, methodological norms, and thematic priorities of subsequent research, collectively constituting the intellectual core of a discipline. The 2015-2025 window ensures analytical coherence by capturing the contemporary phase of ESD development, characterized by procedural maturation and expanding organ-specific applications. The threshold of 1,000 articles was chosen to balance corpus richness—sufficient for statistically robust topic modeling and word-embedding analysis—against analytical tractability, as excessively large, heterogeneous corpora risk introducing noise that may obscure latent structural patterns. This citation-based delimitation strategy is consistent with established practice in computational literature analyses. Inclusion criteria were defined as follows: the study title must contain "endoscopic submucosal dissection," publications must be in English, must be research articles, must be published in citation-indexed journals (SCI, SSCI, SCI-Expanded, ESCI, or SCOPUS), and must be classified under health and/or medicine-related categories within WoS.

Step 2: Data Preprocessing

Data preprocessing was performed to improve the performance of subsequent NLP operations. In this context, the preprocessing procedure began with loading the data into the Google Colab (21) environment. Subsequently, all content in the dataset was converted to lowercase, thereby standardizing letter-case usage. Punctuation marks present in the data that do not provide any semantic contribution were systematically removed. All numerical characters found in the data were cleaned, assuming that numbers were unnecessary for data or semantic analysis. Additionally, common English stop words (e.g., "the," "is," and "a"), which are frequently used in text data but contribute little semantic content, were removed. In the final stage, the data were tokenized into a discrete word list that serves as the fundamental building block for subsequent analyses.

Step 3: Word Embedding Generation

Text data was transformed into dense numerical vector representations to more successfully reveal semantic relationships during the word embedding process (22). For this purpose, the Word2Vec model (23), commonly used in the literature for word embedding, was employed, and training was performed on the token lists generated for each abstract. The vector dimensionality in the model was set to 100, ensuring that each word was represented by a vector of that dimensionality. The choice of vector dimensionality reflects a balanced approach between semantic analysis and computational efficiency. Thus, for each abstract, a single document vector was generated by averaging the Word2Vec vectors of all the words in the abstract. This approach facilitated subsequent ML tasks by consolidating the semantic information of individual words into a fixed-size document vector.

Step 4: Semantic Clustering

The document embedding array was subjected to semantic clustering using the k-means algorithm (24). The elbow method (25) was used to determine the optimal number of clusters (k). This algorithm was run for k values from 2 to 10; the inertia for each k was calculated and plotted. Thus, the point at which the decrease in the inertia value sharply slowed was identified as the elbow point, and the optimal k value was determined to be 2. With this approach, each abstract was labeled with a

semantic cluster label indicating membership in cluster 0 or 1, based on its proximity to the semantic cluster centers. In the next stage, document embeddings in very high dimensions were reduced to two dimensions using the t-Distributed Stochastic Neighbor Embedding (t-SNE) method (26). Thus, the t-SNE model was initialized using different parameters, including the number of components [2]; reproducibility [42]; perplexity [30], which balances local and global structure; and optimization [100]. The obtained two-dimensional coordinates were visualized.

Step 5: Topic Modeling (LDA)

The purpose of topic modeling is to uncover the latent thematic structure within the dataset (27). For this purpose, a corpus was created from tokenized texts, ensuring that each word was matched with a unique integer identifier. With this dictionary, words appearing in fewer than 5 documents or in more than 50% of documents were filtered and refined, thereby eliminating words that are either excessively sparse or excessively common and had low information value. After this process, a Bag-of-Words corpus (28) was created containing word ID and word frequency pairs for each document. Thus, the prepared corpus was subjected to training with the LDA model (29), and this process was carried out within the framework of values regarding the model's reproducibility [42] and the number of training iterations [10]. Prominent words for each topic were identified. The LDA model calculated the probability distribution over all topics for each abstract and assigned the topic with the highest probability as that abstract's dominant topic. Subsequently, word clouds for each topic were created from words and their weights using the WordCloud method (30) of the WordCloud library.

Step 6: Topic Correlation Analysis

The purpose of topic correlation analysis is to analyze the relationships and co-occurrence patterns between identified latent topics (31). In this context, topic distributions for each document were obtained from the LDA model. These data were organized as a topic matrix, in which rows represent documents and columns represent topic probabilities. Then, Pearson correlation (32) coefficients were calculated for all topic column pairs in the matrix, thus obtaining a topic correlation matrix showing the extent to which different topics co-occur across documents. This correlation matrix was visualized in the form of a heatmap using the heatmap function (33).

Step 7: Temporal Dynamics Analysis

The purpose of analysis of temporal dynamics is to examine how the prevalence of semantic clusters and topics has evolved. Data obtained from topic-matrix dataframe indices were consolidated with the original dataset, and averages of topic probabilities were calculated for each year, producing a dataframe that shows the annual average prevalence of each topic. The obtained time series data were visualized through line graphs using pyplot (34), thereby revealing the time-dependent development of each topic. Similarly, another data frame was created by calculating the proportion of documents in each semantic cluster per year. These data were also visualized using line graphs, which present in detail how the overall semantic structure changed over time and highlight prominent shifts between clusters.

Step 8: Performance Evaluation

The performance of the k-means clustering algorithm was analyzed using two fundamental metrics: the Silhouette score (35) and the Davies-Bouldin index (36). In this regard, the Silhouette score is a relative measure that quantifies an object's similarity to its own cluster relative to its similarity to other clusters. The Davies-Bouldin index evaluates the average similarity ratio of each cluster to its most similar cluster. In the topic modeling process, the interpretability of LDA topics was evaluated using the coherence score. Topic coherence is widely considered a sufficient and reliable metric for evaluating the quality of topics generated by LDA models, as it reflects the semantic interpretability of topics in a manner consistent with human judgment. Coherence measures have been shown to correlate strongly with human assessments of topic quality, supporting their use as a primary evaluation criterion in topic modeling (37).

Statistical Analysis

All computational analyses were performed using Google Colaboratory (21). Python (38) was used as the primary programming language for the analysis. The computational workflow was supported by various libraries commonly used in scientific research. These include NumPy (39) for numerical operations; Pandas (40) for data processing and tabular analyses; Matplotlib (34) and Seaborn (33) for visualization; Scikit-learn (41) for ML operations; Statsmodels (42) for time series modeling; and WordCloud (43) for word-cloud visualizations.

Results

Semantic Clustering Analysis Findings

Findings related to semantic clustering analysis are shown in Figure 2.

The blue points in the graph are predominantly concentrated on the right side, particularly in the lower-right region, while the green points are predominantly located on the left side, especially in the upper-left region. Although the clusters are distinct, minimal overlap between points of the two clusters is observable in the middle section.

Topic Modeling Findings

The LDA model identified 9 main topics within the corpus, with a coherence score of 0.39. The results are summarized in Table 1.

Findings related to word clouds are presented in Figure 3.

Concepts such as "patient," "ESD," "endoscopic," "resection," "bleeding," "dissection," "gastric," and "colorectal" are encountered in nearly all topics. While the words "patient," "ESD," and "group" stand out in Topic 0, the words "bleeding" and "patients" are prominent in Topic 2. For Topics 3 and 5, the concepts of "submucosal" and "resection" are prominent. In Topic 4, the word "perforation" stands out.

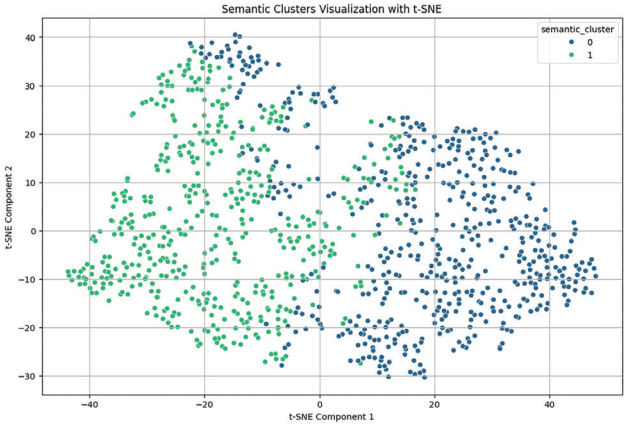

Figure 2. Two-dimensional visualization of semantic clusters generated using t-distributed stochastic neighbor embedding (t-SNE). Each point represents one publication abstract, while colors indicate cluster assignment derived from K-means clustering. Spatial proximity between points reflects semantic similarity among publications

Findings on Topic Correlations

The heatmap showing correlations between topics is presented in Figure 4a.

Correlation analysis revealed that all inter-topic associations were negative, ranging from mild to moderate (-0.22 to -0.04). In this regard, the strongest negative correlations were observed between Topic 0 and Topic 2 (-0.21), Topic 3 (-0.22), and Topic 7 (-0.22), suggesting that documents heavily weighted toward Topic 0, centered on technical innovation and procedural optimization, are least likely to concurrently emphasize the thematic content represented by these three topics.

Findings on Temporal Dynamics Analysis

Time-dependent changes in clusters and topics are presented in Figure 4b.

Topics exhibited a relatively stable distribution over the years examined. While most topics fluctuate, ranging from 5% to 30% in annual document proportions, no dramatic change in topics was observed. However, Topic 7 showed a marked increase during the examined period, reaching approximately 75%.

The right graph (Cluster 0-blue; Cluster 1-orange) showed that, between 2015 and 2023, both clusters exhibited fluctuations at similar rates, with values generally ranging between 40% and 60%. Cluster 0 was more dominant, exceeding 60% in 2017 and 2022, while Cluster 1 remained at relatively lower levels during these periods. In the subsequent period (2024-2025), Cluster 0 showed a sharp increase exceeding 75%, while Cluster 1 experienced a dramatic decline to below 25%.

Table 1. Topics and prominent keywords

Topic 0	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8
esd	group	patients	esd	esd	esd	patients	esd	esd
patients	esd	bleeding	submucosal	patients	resection	esd	patients	bleeding
group	patients	esd	endoscopic	pecs	endoscopic	risk	resection	patients
resection	p	delayed	resection	perforation	dissection	factors	lesions	esophageal
p	esophageal	closure	dissection	p	submucosal	submucosal	endoscopic	stricture
endoscopic	gastric	group	cases	endoscopic	p	lesions	p	endoscopic
survival	endoscopic	endoscopic	procedure	risk	time	gastric	group	submucosal
recurrence	stricture	p	lesions	submucosal	speed	invasion	vs	lesions
outcomes	vs	risk	patients	factors	emr	cancer	colorectal	dissection
submucosal	ulcer	rate	study	study	procedure	ci	submucosal	gastric
rate	dissection	mucosal	colorectal	dissection	traction	endoscopic	rate	group
treatment	groups	submucosal	time	colorectal	cesd	lnm	study	p
gastric	study	colorectal	gastric	syndrome	rate	egc	treatment	postesd
dissection	significantly	study	performed	analysis	using	tumor	dissection	delayed
cancer	submucosal	gastric	p	group	study	fibrosis	bloc	risk

Figure 3. Word clouds representing the dominant keywords within each LDA-derived topic. Word size is proportional to the relative importance (weight) of each term within the corresponding topic
LDA: Latent Dirichlet Allocation

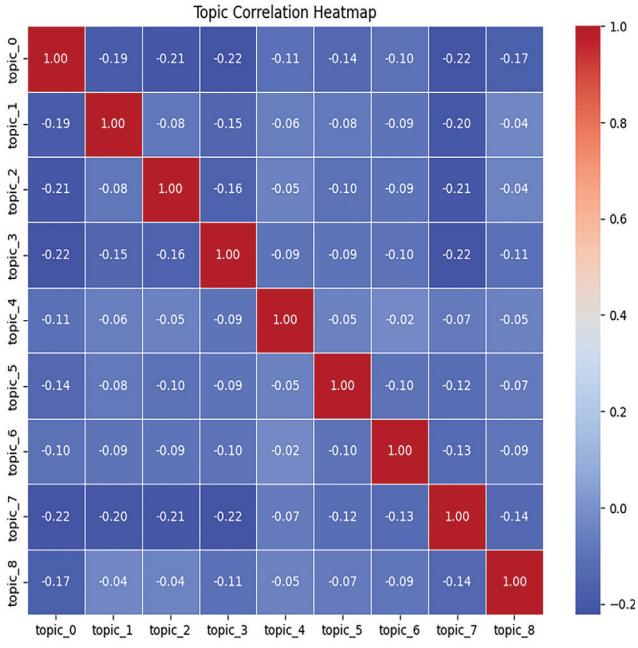

Figure 4a. Heatmap illustrating Pearson correlation coefficients between LDA-derived topics. Warmer and cooler colors represent the strength and direction of correlations, indicating the degree of thematic overlap or separation between topics
LDA: Latent Dirichlet Allocation

Findings on Cluster-Topic Distribution

Findings on cluster and topic distribution are presented in Figure 5a.

Accordingly, a balanced distribution between the two clusters was observed during the 2015-2023 period. Although both clusters had similar values of approximately 60 in 2017, Cluster 0 stood out in 2018. However, Cluster 1 lagged behind Cluster 0 in this regard. Subsequently, similar fluctuations were observed between 2019 and 2023, and, generally, no major differences were observed between the two clusters.

The right-hand graph shows values corresponding to the average probability distribution of topics (0-8) within two semantic clusters. Topic 0 (0.29) and Topic 7 (0.31) stood out in Cluster 0. Other topics had low probabilities in Cluster 0 (between 0.01 and 0.11).

Findings on Model Performance

Performance graphs for the semantic clustering and topic modeling processes are presented in Figure 5b.

The highest Silhouette score was obtained with two semantic clusters and was approximately 0.215. The right-hand graph presents the coherence score for LDA topic modeling. Findings from the LDA model show that nine topics have the highest coherence scores, suggesting that the model achieves a satisfactory level of semantic coherence and that the resulting topics are meaningful and interpretable for further analysis.

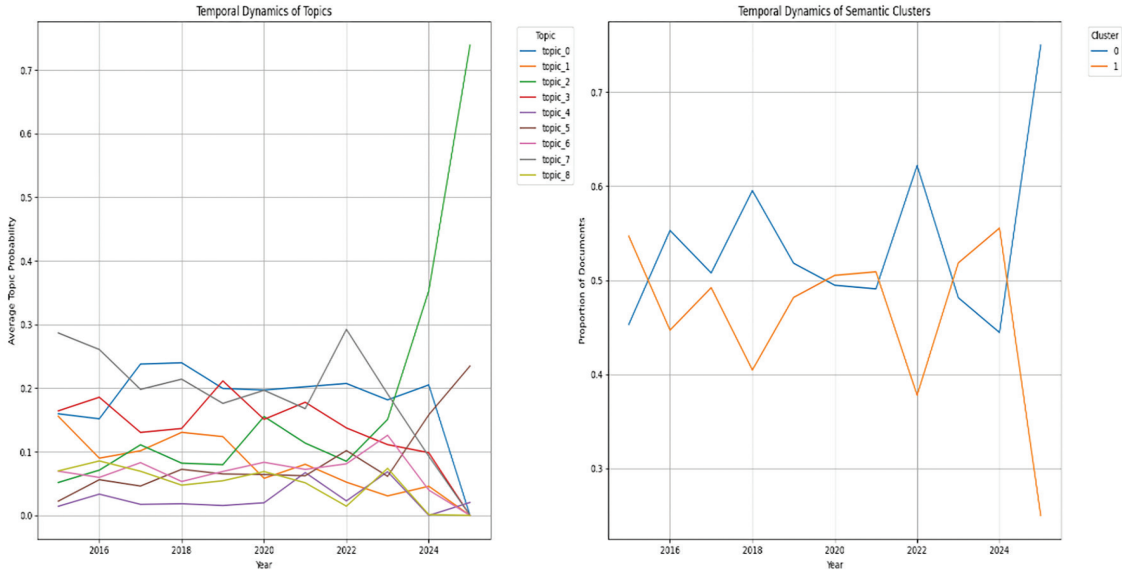

Figure 4b. Annual temporal evolution of semantic clusters and LDA-derived topics between 2015 and 2025. The graphs illustrate changes in the relative prevalence of semantic clusters and thematic topics over time, highlighting shifts in research focus within the ESD literature

LDA: Latent Dirichlet Allocation, ESD: Endoscopic submucosal dissection

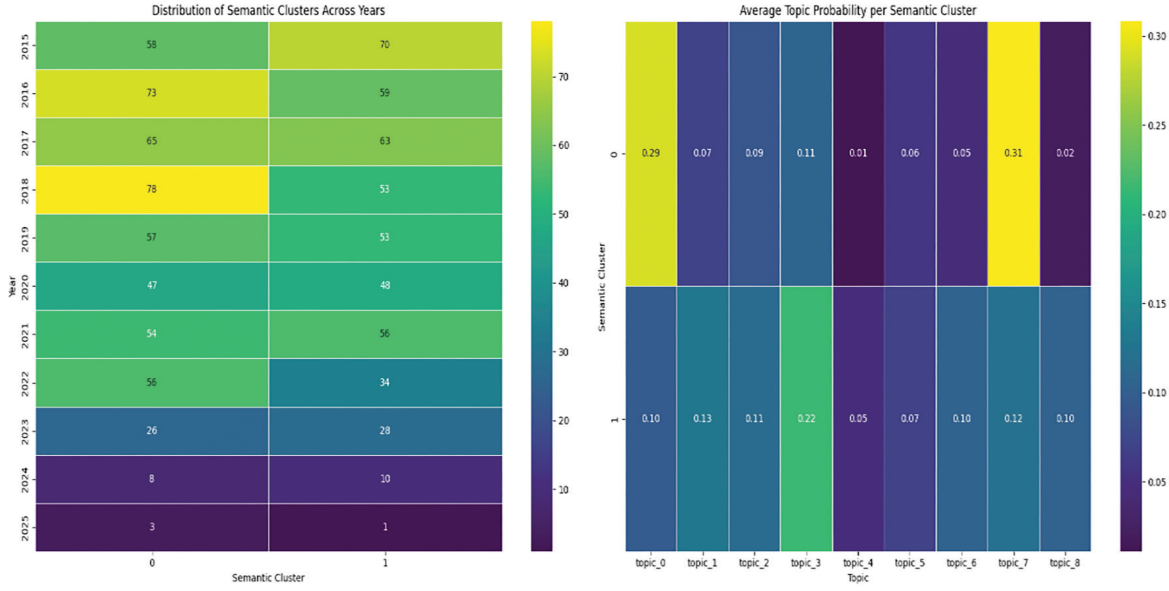

Figure 5a. Distribution of semantic clusters over time and average topic probabilities within each semantic cluster. The left panel illustrates annual changes in semantic cluster distribution, whereas the right panel presents the mean probability of each LDA-derived topic within Cluster 0 and Cluster 1

LDA: Latent Dirichlet Allocation

Discussion

This study represents one of the first comprehensive analyses using NLP and ML methods to reveal the semantic and thematic structure of the ESD literature. Since its development in Japan in the late 1990s (1) to overcome the limitations of EMR (2,3), ESD has emerged as a significant paradigm shift in therapeutic endoscopy. Our findings demonstrate that the literature surrounding the procedure is organized around two main axes: technical innovation and procedural development, and clinical outcomes and organ-specific applications.

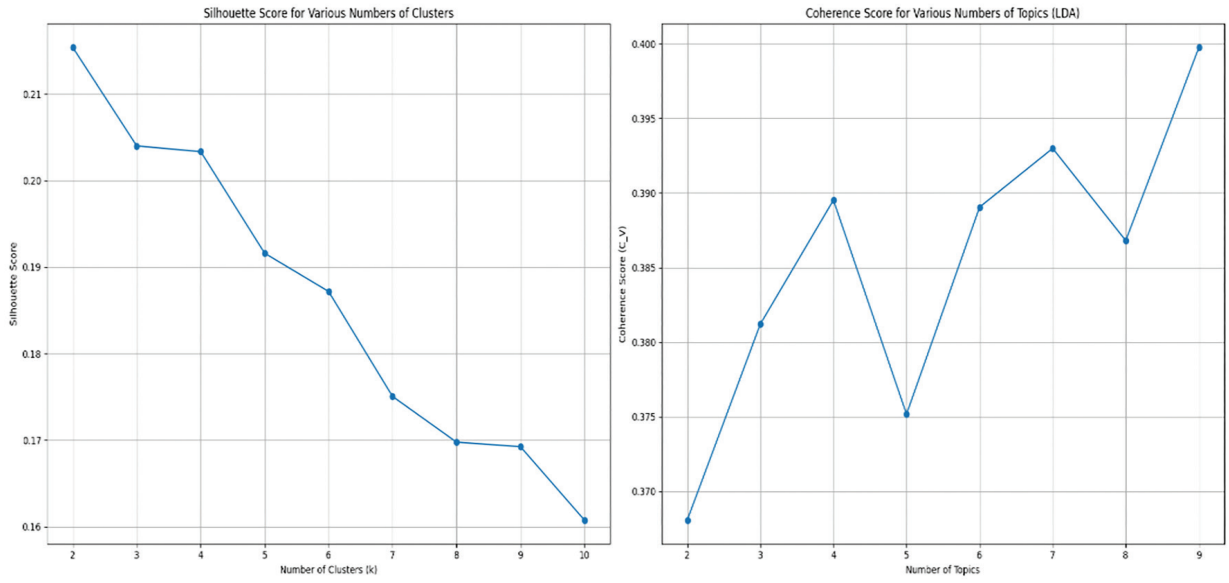

Figure 5b. Performance evaluation of semantic clustering and topic modeling. The left panel presents silhouette scores across candidate cluster numbers, whereas the right panel illustrates coherence scores for different LDA topic models
LDA: Latent Dirichlet Allocation

The semantic clustering analysis revealed that the ESD literature exhibits a fundamental structural divergence. This bifurcation was consistently observed in both t-SNE visualizations and LDA-based topic modeling, revealing two parallel research streams: clinical and technical. Studies related to clinical outcomes and organ-specific applications are positioned together in semantic space, whereas themes of dissection methods, device development, and procedural efficiency constitute a separate, technical research area. This divergence indicates that the causal relationships between technical innovation and clinical performance are not addressed holistically in the literature. K-means clustering was performed with k=2, resulting in a low silhouette score of 0.21. This relatively low value indicates limited separation between clusters, suggesting that the chosen number of clusters may oversimplify the underlying structure of the data. While the two-cluster solution provides a basic overview of the dataset, it may fail to capture more nuanced patterns that could be revealed with alternative clustering strategies or a higher number of clusters. Future analyses should consider testing additional k values and validating cluster stability using complementary metrics to ensure that the identified groupings accurately reflect the complexity of the dataset. These limitations highlight the need for cautious interpretation of the clustering results and underscore that the findings should be considered exploratory rather than definitive.

This structural separation is particularly significant given that ESD's clinical success fundamentally depends on technical proficiency. The technique's ability to achieve en bloc resection of lesions exceeding 20 mm—where conventional EMR requires piecemeal resection, with associated high local recurrence rates (7)—depends entirely on the operator's technical skill and the available instrumentation. The procedure involves submucosal lifting with viscous agents such as sodium hyaluronate followed by precise dissection with specialized knives (11), and the reported complication incidences of 3.5%, 3.3%, and 4.6% for gastric, esophageal, and colorectal ESD, respectively (12), underscore the intimate connection between technical execution and clinical outcomes. Yet our analysis reveals that these two dimensions are largely explored in isolation in the literature. This situation highlights that clinical needs are not taken as the primary basis for health innovation; technological advancements remain largely disconnected from the problems encountered in healthcare; consequently, the convergence between clinical practice and technology is becoming increasingly difficult.

The lack of clinical-technical integration emerges as an important limiting factor in the development of ESD and suggests that research designs need to be structured with a more holistic approach. This gap is particularly critical given that factors such as submucosal fibrosis caused by prior interventions like biopsy or marking can increase dissection difficulty and reduce en bloc success rates (13), creating a direct pathway from technical challenges to clinical outcomes. The semantically weak connections among complication-focused themes further reveal that

risk prediction models are still immature, despite the procedure's well-documented association with bleeding and perforation risks. Accordingly, there is a clear need in the literature for modeling studies aimed at predicting certain risks, such as delayed bleeding.

The temporal analysis revealed a marked increase in esophageal and gastric studies after 2024, whereas colorectal ESD research remained relatively stagnant. This thematic shift provides empirical support for the diffusion of innovations model (19), which explains the adoption of health technologies through distinct stages. The observed pattern reflects the natural progression from technical feasibility studies to widespread clinical implementation.

Early-stage ESD research focusing on technical feasibility and instrument innovation corresponds to the "knowledge acquisition" and "persuasion" stages of the innovation adoption process (19). The current concentration of publications on upper gastrointestinal applications, particularly the acceptance of ESD as the standard treatment for superficial gastric neoplasias and early-stage gastric cancer (2), as well as its increasing adoption as first-line treatment for superficial esophageal squamous cell carcinoma, indicates progression to the "implementation" and "confirmation" stages. Applications carried out in more limited centers during the technical feasibility and early-experience phases have expanded rapidly upon reaching a certain maturity, thereby creating thematic density in the literature. In this context, the evidence suggests that the next phase of ESD research should move beyond adoption-focused discourse toward risk stratification and outcome optimization, thereby reinforcing the transition from widespread implementation to a more data-driven and precision-oriented confirmation stage in clinical practice.

The concentration of publications in organ-specific areas is also consistent with learning-curve dynamics. The relative stagnation in colorectal ESD research, despite EMR remaining the dominant method in this region (8), may reflect the unique technical challenges posed by this anatomical site. The distal colon and rectum present specific difficulties related to wall thickness, peristalsis, and anatomical configuration, which may have slowed the adoption curve relative to upper gastrointestinal applications. This finding suggests that colorectal ESD may still be in an early phase of the diffusion process, potentially transitioning between the "persuasion" and "decision" stages. With the adoption of ESD in this field, the accumulation of experience and knowledge is expected to yield substantially greater benefits for clinical practice.

This study goes beyond classical bibliometric approaches and presents an innovative framework that objectively reveals the structural characteristics of the literature through word embedding-based semantic analysis. LDA-based topic modeling (18) made the thematic centers of the literature quantitatively visible, while t-SNE-based visualization enabled better understanding of high-dimensional semantic relationships. The combined use of these methods allowed the ESD literature to be evaluated from a structural perspective, revealing patterns that cannot be detected through traditional systematic reviews (18).

These findings are critical to guiding clinical practices and research priorities. The marked increase in publications on the upper gastrointestinal system demonstrates that clinical practice increasingly focuses on these areas and that training programs need to allocate more resources to esophageal and gastric dissection techniques. The scarcity of sufficiently experienced operators in Western countries, which has limited widespread adoption, can be addressed through structured training programs (14,15) informed by the learning-curve dynamics revealed in our analysis.

The structural divergence between technical and clinical research reveals that procedural innovation is not evaluated in an integrated manner alongside clinical outcomes. This situation demonstrates the need for more holistic research in critical areas such as management of complications, selection of dissection strategies, and effectiveness of device use. Given that ESD provides superior outcomes in specific clinical contexts (8)—particularly in lesions exceeding 20 mm, cases with suspected superficial submucosal invasion, or cases unsuitable for conventional snare polypectomy, especially in the distal colon and rectum (9,10)—the integration of technical and clinical research streams could optimize patient selection and procedural planning.

The technique's ability to obtain intact specimens provides histopathological advantages by enabling clear assessment of lateral and vertical surgical margins and accurate analysis of invasion depth (4,5). Such diagnostic precision both enhances the success of curative treatment and reduces the risk of recurrence; however, our analysis suggests that the relationship between the quality of technical execution and these histopathological outcomes remains underexplored in the literature.

A fixed corpus size of 1,000 publications, selected on a citation-priority basis, constitutes a notable limitation of this study. This threshold, while methodologically convenient, is analytically arbitrary: it imposes a rigid ceiling that automatically excludes any publication—however recent, innovative, or clinically significant—that falls outside the citation-ranked top tier at the time of data extraction. In a rapidly evolving procedural field such as ESD, where AI-based risk prediction, organ-specific protocol refinement, and novel technical variants are actively emerging, this cutoff systematically filters out the studies most likely to signal where the field is heading rather than where it has

been. The resulting corpus is therefore best understood not as a representative sample of ESD research, but as a citation-weighted retrospective of its most consolidated subset. Future bibliometric and NLP-based investigations should consider recency-stratified or citation-agnostic sampling strategies to counterbalance this structural skew and more accurately capture the field's emerging frontier.

Study Limitations

This study has several limitations that should be acknowledged. Most critically, the analysis is restricted to abstract-level text, which constitutes a significant methodological constraint. Abstracts inherently omit procedural details, statistical nuances, subgroup-level findings, and contextual clinical information embedded in full-text publications, which may carry substantial thematic weight. This restriction likely introduces systematic bias into the semantic and topical structures identified, as abstracts selectively emphasize outcomes over methodology and context. Consequently, the thematic architecture revealed in this study should be interpreted as a representation of how ESD research is framed and communicated, rather than a complete reflection of its intellectual depth.

A second and equally consequential limitation concerns selection bias introduced by the citation-based corpus construction strategy. By restricting the analysis to the top 1,000 most-cited publications, this study systematically privileges well-established, highly cited work to the detriment of recent contributions that have not yet had sufficient time to accumulate citations. Given that citation accumulation is inherently a slow, cumulative process—often spanning several years post-publication—this sampling approach almost certainly underrepresents studies published toward the latter portion of the 2015–2025 window, including emerging technical innovations, novel AI-based applications, and early-stage organ-specific protocols, which may not yet be reflected in the citation record. As a direct consequence, the thematic and temporal structures identified in this study should be understood to characterize the established, citation-validated core of ESD research rather than its full, current state; nascent research trajectories, cutting-edge methodological developments, and recently emerging subfields are likely underrepresented or entirely absent from the derived topic structure. This limitation is particularly consequential for the temporal trend analysis, as citation-driven undersampling of recent publications may artificially compress or obscure both the acceleration in upper-gastrointestinal and AI-related research that more recent, lower-citation studies might otherwise reveal. Finally, the 2015-2025 analytical window, which is temporally defined, may be further influenced by publication delays, geographic disparities in research output, and field-specific innovation cycles—factors that interact with and potentially compound the citation-based selection bias described above.

Notwithstanding these limitations, this study's primary strength lies in its novel application of NLP and unsupervised machine learning to the ESD literature, offering an objective, data-driven alternative to traditional narrative or systematic reviews. Word embedding-based semantic analysis captures nuanced conceptual relationships between research constructs that conventional keyword-based approaches fail to detect, while LDA topic modeling systematically exposes latent thematic clusters and structural divergences across the corpus. The integration of temporal analysis further enabled the longitudinal mapping of research trajectories—a dimension largely absent from existing ESD literature reviews. Collectively, these methodological layers render the study's findings reproducible, scalable, and transferable to other domains of procedural medicine seeking macro-structural characterization of their research landscapes.

Future ESD research should prioritize four key areas that together define the field's methodological and clinical trajectory. First, the current literature inadequately bridges technical innovation and clinical outcomes; multidisciplinary studies that jointly evaluate both dimensions within unified analytical frameworks are urgently needed to close this gap. Second, organ-specific comparative research—spanning esophageal, gastric, and colonic ESD—remains scarce despite the markedly distinct difficulty levels and complication profiles that characterize each anatomical site. Third, AI-based clinical decision support models capable of real-time prediction of complication risks, such as perforation and bleeding, represent a critical yet underdeveloped frontier, reflecting the broader immaturity of current risk-prediction frameworks in this domain. Fourth, and perhaps most pressing for global access to the procedure, quantitative modeling of ESD learning curves and standardized, competency-based performance metrics are essential to guide training programs, particularly in Western countries, where the scarcity of operators continues to constrain the expansion of this technique. Taken together, these four priorities—integrated technical-clinical evaluation, organ-specific comparative modeling, AI-driven adverse event prediction, and standardized competency assessment—chart a coherent agenda for advancing both methodological rigor and clinical relevance of ESD research.

Conclusion

This study provides a systematic mapping of the semantic and thematic architecture of the ESD literature using computational text analytics, extending beyond conventional bibliometric approaches. The analysis identifies two structurally distinct thematic clusters within

ESD research: one centered on technical innovation and procedural optimization, and the other on clinical outcomes and organ-specific therapeutic protocols. Co-occurrence and topic-relationship analyses show limited overlap between these two clusters, indicating that technical and clinical themes are predominantly addressed in separate bodies of literature rather than within integrated studies. Temporal analysis demonstrates a measurable shift in publication focus toward upper gastrointestinal applications, particularly esophageal and gastric ESD, in recent years. This shift is consistent with a broader transition observed in the data, from early-stage feasibility studies toward organ-specific clinical implementation. Future studies integrating transformer-based language models and full-text semantic analysis may further improve the understanding of knowledge evolution in ESD research.

Ethics

Ethics Committee Approval: This study was based exclusively on publicly available publication data retrieved from bibliometric databases. The analysis did not involve any human subjects, primary data collection, or interventions. Since all data used in this research were already in the public domain and freely accessible, and no identifiable personal information or sensitive data were processed, ethical approval from an institutional review board was not required for this bibliometric analysis.

Informed Consent: Informed consent was not applicable to this study, as it involved a bibliometric analysis of publicly available bibliographic data and did not involve human participants or the collection of identifiable personal information.

Footnotes

Authorship Contributions

Concept: H.D., T.A., S.M., Design: H.D., M.T., Data Collection or Processing: H.D., M.T., Analysis or Interpretation: H.D., M.T., Literature Search: H.D., T.A., S.M., M.T., Writing: H.D., M.T.

Conflict of Interest: The authors declared that there were no conflicts of interest.

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

Declaration of AI Use: The article was written solely by the authors. OpenAI's ChatGPT model was used to translate the manuscript into English, and Anthropic's Claude AI model assisted in reviewing the translated text to ensure linguistic accuracy and professional quality.

References

1. Fukuzawa M, Gotoda T. History of endoscopic submucosal dissection and role for colorectal endoscopic submucosal dissection: A Japanese perspective. Gastrointest Interv. 2012;1:30-5.
2. Moutzoukis MK, Manolakis A, Kapsoritakis A, Christodoulou D, Argyriou K. Endoscopic submucosal dissection: Challenges, innovations, and the road ahead. World J Gastrointest Endosc. 2026;18:113918.
3. Elkholy S, Awad A, Haggag H, Essam K, Yousef K, El Sherbiny M. Different endoscopic submucosal dissection techniques: a tailored approach. VideoGIE. 2025;10:277-84.
4. Pal S, Bhaduri G. Endoscopic submucosal dissection for early gastrointestinal malignancies: Current state and future perspectives. World J Gastrointest Endosc. 2025;17:109144.
5. Ziachehabi A, Worm M, Liu DHW, Pimingstorfer P, Langer R. Endoscopic submucosal dissection (ESD) of upper gastrointestinal carcinomas: an integrated clinical and pathological perspective. J Clin Med. 2025;14:8817.
6. Zheng X, Xu L. Endoscopic submucosal dissection vs. endoscopic mucosal resection for early gastric cancer: a meta-analysis. Front Med (Lausanne). 2025;12:1702512.
7. Bourke MJ, Neuhaus H, Bergman JJ. Endoscopic submucosal dissection: Indications and application in Western endoscopy practice. Gastroenterology. 2018;154:1887-900.
8. Matsui N, Akahoshi K, Nakamura K, Ihara E, Kita H. Endoscopic submucosal dissection for removal of superficial gastrointestinal neoplasms: a technical review. World J Gastrointest Endosc. 2012;4:123-36.
9. Mendes RR, Barreiro P, Mascarenhas A, Franco AR, Carvalho L, Chagas C. Endoscopic submucosal dissection for resections larger than 10 cm: outcomes from a Portuguese center. GE Port J Gastroenterol. 2023;31:33-40.
10. Bronswijk M, Rasschaert G, Hayashi Y, Yamamoto H. Colorectal endoscopic submucosal dissection: outcomes from a European multicenter cohort. Endoscopy. 2023;55:130-9.
11. Yamamoto H. Endoscopic submucosal dissection of early cancers and large flat adenomas. Clin Gastroenterol Hepatol. 2005;3:S74-6.
12. Odagiri H, Yasunaga H. Complications following endoscopic submucosal dissection for gastric, esophageal, and colorectal cancer. Ann Transl Med. 2017;5:189.
13. Kuroha M, Shiga H, Kanazawa Y, et al. Factors associated with fibrosis during colorectal endoscopic submucosal dissection. Digestion. 2021;102:590-8.
14. Bhatt YM, Weilert F, Fredrick RT, et al. EUS-guided treatment of gastric fundal varices with combined injection of coils and cyanoacrylate glue: A large U.S. experience over 6 years. Gastrointest Endosc. 2016;83:1164-72.

15. Barbour JA, O'Toole P, Suzuki N, Dolwani S. Learning endoscopic submucosal dissection in the UK: Barriers, solutions and pathways for training. Frontline Gastroenterol. 2020;12:671-6.
16. Akintoye E, Kumar N, Obaitan I, Alayo QA, Thompson CC. Peroral endoscopic myotomy: a meta-analysis. Endoscopy. 2016;48:1059-68.
17. Hatta W, Ogata Y, Ogasawara K, et al. Recent advances in endoscopic submucosal dissection for gastric cancer: focusing on expanded indications and technological innovations. J Gastric Cancer. 2026;26:76-91.
18. Vajjala S, Majumder B, Gupta A, Surana H. Practical natural language processing: A comprehensive guide to building real-world NLP systems. O'Reilly Media; 2020.
19. Rogers EM. Diffusion of innovations. 5th ed. New York: Free Press; 2003.
20. Li K, Rollins J, Yan E. Web of Science use in published research and review papers 1997-2017: a selective, dynamic, cross-domain, content-based analysis. Scientometrics. 2018;115:1-20.
21. Google. Google Colaboratory [cloud computing platform]. 2025. Available from: https://colab.research.google.com
22. Meddeb A, Ben Romdhane L. Using topic modeling and word embedding for topic extraction in Twitter. Procedia Comput Sci. 2022;207:790-9.
23. Mikolov T, Chen K, Corrado G, Dean J. Efficient estimation of word representations in vector space. arXiv. 2013.
24. Lloyd SP. Least squares quantization in PCM. IEEE Trans Inf Theory. 1982;28:129-37.
25. Ketchen DJ Jr, Shook CL. The application of cluster analysis in strategic management research: an analysis and critique. Strateg Manag J. 1996;17:441-58.
26. van der Maaten L, Hinton G. Visualizing data using t-SNE. J Mach Learn Res. 2008;9:2579-605.
27. Blei DM, Carin L, Dunson D. Probabilistic topic models. IEEE Signal Process Mag. 2010;27:55-65.
28. Manning CD, Raghavan P, Schütze H. Introduction to information retrieval. Cambridge: Cambridge University Press; 2008.
29. Blei DM, Ng AY, Jordan MI. Latent Dirichlet allocation. J Mach Learn Res. 2003;3:993-1022.
30. Skeppstedt M, Ahltorp M, Kucher K, Lindström M. From word clouds to Word Rain: Revisiting the classic word cloud to visualize climate change texts. Inf Vis. 2024;23:217-38.
31. Blei DM, Lafferty JD. Correlated topic models. Adv Neural Inf Process Syst. 2006;18:147-54.
32. Pearson K. Note on regression and inheritance in the case of two parents. Proc R Soc Lond. 1895;58:240-2.
33. Waskom ML. Seaborn: Statistical data visualization. J Open Source Softw. 2021;6:3021.
34. Hunter JD. Matplotlib: A 2D graphics environment. Comput Sci Eng. 2007;9:90-5.
35. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53-65.
36. Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell. 1979;1:224-7.
37. Mimno D, Wallach HM, Talley E, Leenders M, McCallum A. Optimizing semantic coherence in topic models. In: Barzilay R, Johnson M, editors. Proc 2011 Conf Empirical Methods Nat Lang Process. Association for Computational Linguistics; 2011.
38. Python Software Foundation. Python (Version 3.9.18) [programming language]. 2024. Available from: https://www.python.org/
39. Harris CR, Millman KJ, van der Walt SJ, et al. Array programming with NumPy. Nature. 2020;585:357-62.
40. McKinney W. Data structures for statistical computing in Python. In: Proc 9th Python Sci Conf; 2010.
41. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825-30.
42. Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with Python. Proceedings of the 9th Python in Science Conference. 2010:92-6.
43. Mueller A. WordCloud: a little word cloud generator in Python. Python Package Index (PyPI). 2023.

Original Article

DOI: 10.4274/haseki.galenos.2026.96729
Med Bull Haseki 2026;64(4):345-353

Real-world Efficacy, Safety, and Drug Survival of IL-23 Inhibitors in Overweight and Obese Patients with Psoriasis: A 52-week Retrospective Study

Gizem Pehlivan Ulutas, Yavuz Semiz

University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Dermatology and Venereology, Istanbul, Türkiye

Abstract

Aim: Obesity is a common comorbidity in psoriasis and has been associated with reduced response to biologic therapies. However, data on the real-world effectiveness of interleukin (IL)-23 inhibitors in overweight and obese patients with psoriasis remain limited. This study aimed to evaluate the efficacy, safety, and drug survival of IL-23 inhibitors in this patient population over 52 weeks.

Methods: This retrospective, single-center observational study included patients with psoriasis and a body mass index (BMI) ≥25 kg/m^2 who were treated with guselkumab or risankizumab between November 2021 and November 2025. Efficacy was assessed using absolute Psoriasis Area and Severity Index (PASI) scores and PASI75/90/100 response rates at predefined time points. Drug survival, adverse events, and reasons for discontinuation were evaluated. Multivariable logistic regression analysis was performed to identify independent predictors of achieving PASI90 at week 52.

Results: Among the 75 included patients, the mean PASI score at baseline was 13.7±7.7, which improved to 0.8±1.1 at week 52. Psoriasis Area and Severity Index 90 and PASI100 response rates increased from 37.3% and 12.0% at weeks 12-16 to 77.5% and 56.3% at week 52, respectively. Overweight patients achieved significantly higher PASI90 response rates than obese patients at weeks 28 and 52. In multivariable analysis, lower BMI and biologic-naïve status remained independent predictors of achieving PASI90 at week 52. The 52-week drug survival rate was 94.7%. Treatment was discontinued in 4 patients (5.3%), most commonly because of secondary loss of response. The most common adverse events were infections. No cases of hepatitis B or tuberculosis reactivation were observed.

Conclusion: Interleukin-23 inhibitors demonstrated sustained clinical effectiveness, high 52-week drug survival, and a favorable safety profile in overweight and obese patients with psoriasis. Although lower BMI was independently associated with higher odds of achieving PASI90 by week 52, drug survival remained comparable across BMI groups, supporting the use of IL-23 inhibitors in clinical practice.

Keywords: Psoriasis, guselkumab, risankizumab, interleukin-23, body mass index, treatment outcome

Introduction

Psoriasis is a chronic inflammatory disease affecting approximately 2-3% of the global population and around 1% in Türkiye (1,2). Previously considered a skin-limited condition, it is now widely accepted as a systemic disorder with multiple comorbidities, including obesity, type 2 diabetes mellitus, dyslipidemia, atherosclerotic cardiovascular disease, and metabolic-associated liver disease (3,4). A 1.5 times higher risk of obesity is seen in the psoriasis population relative to control subjects, marking it as one of the most widespread concurrent conditions (5). Moreover, obesity is associated with greater disease severity and an increased prevalence of psoriatic arthritis (5,6).

Biologics have significantly improved the management of psoriasis by targeting key inflammatory pathways and providing high levels of efficacy and durable disease control (7). However, obesity may influence the pharmacokinetics of biologic agents by altering drug distribution and clearance, potentially leading to

Corresponding Author: Gizem Pehlivan Ulutas, MD, University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Dermatology and Venereology, Istanbul, Türkiye
E-mail: gizempehlivan93@gmail.com **ORCID:** orcid.org/0000-0002-5906-3976
Received: 15.05.2026 **Accepted:** 01.08.2026 **Publication Date:** 24.09.2026

Cite this article as: Pehlivan Ulutas G, Semiz Y. Real-world efficacy, safety, and drug survival of IL-23 inhibitors in overweight and obese patients with Psoriasis: a 52-week retrospective study. Med Bull Haseki. 2026;64(4):345-353

reduced treatment efficacy (8). While some biologics allow weight-based dosing, most are administered at fixed doses. Reduced treatment responses have been reported with tumor necrosis factor-α (TNF-α) inhibitors in obese patients (9,10). According to data derived from a large multicenter registry, obesity is associated with a 25% and 30% reduction in PASI75 and PASI90 response rates, respectively, among patients receiving TNF-α and interleukin (IL)-23/Th17-targeted biologics (11). Notably, this effect appeared more pronounced in patients treated with TNF-α and IL-17 inhibitors. In line with these findings, higher body mass index (BMI) has also been identified as a predictor of shorter drug survival, largely driven by reduced treatment effectiveness (12). Despite the high efficacy of guselkumab and risankizumab in the overall psoriasis population, data on their performance in overweight or obese patients remain limited. We hypothesized that IL-23 inhibitors would provide sustained clinical effectiveness, a favorable safety profile, and high drug survival in overweight and obese patients with psoriasis.

Based on this hypothesis, the present study aimed to evaluate the effectiveness, safety, and drug survival of IL-23 inhibitors in overweight and obese patients with psoriasis over 52 weeks of follow-up. By providing real-world data from this patient population, this study may contribute to optimizing treatment strategies and supporting evidence-based therapeutic decision-making in routine clinical practice.

Materials and Methods

Compliance With Ethical Standards

The study received approval from the University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital Non-Interventional Clinical Research Ethics Committee (approval number: 106-2026, date: 22.04.2026) and was conducted in accordance with the principles of the Declaration of Helsinki. Informed consent was waived due to the retrospective design of the study. All data were anonymized, and no identifiable patient information was used.

Study Design

This retrospective, single-center observational study was conducted at a tertiary referral center with a dedicated psoriasis outpatient clinic. Medical records of patients treated with biologics from November 2021 to November 2025 were reviewed. The patient selection process is summarized in Figure 1. Adult patients (≥18 years) with plaque psoriasis who received IL-23 inhibitors (guselkumab or risankizumab) and were classified as overweight (BMI 25-29.9 kg/m^2) or obese (BMI ≥30 kg/m^2) were included. Patients with a BMI <25 kg/m^2, those treated with biologics other than IL-23 inhibitors, and those with incomplete baseline or follow-up data were excluded. The primary study outcomes were absolute PASI scores and PASI75, PASI90, and PASI100 response rates at weeks 4, 12-16, 28, and 52. Secondary outcomes included drug survival; treatment discontinuation rates and reasons for discontinuation; and the occurrence of adverse events.

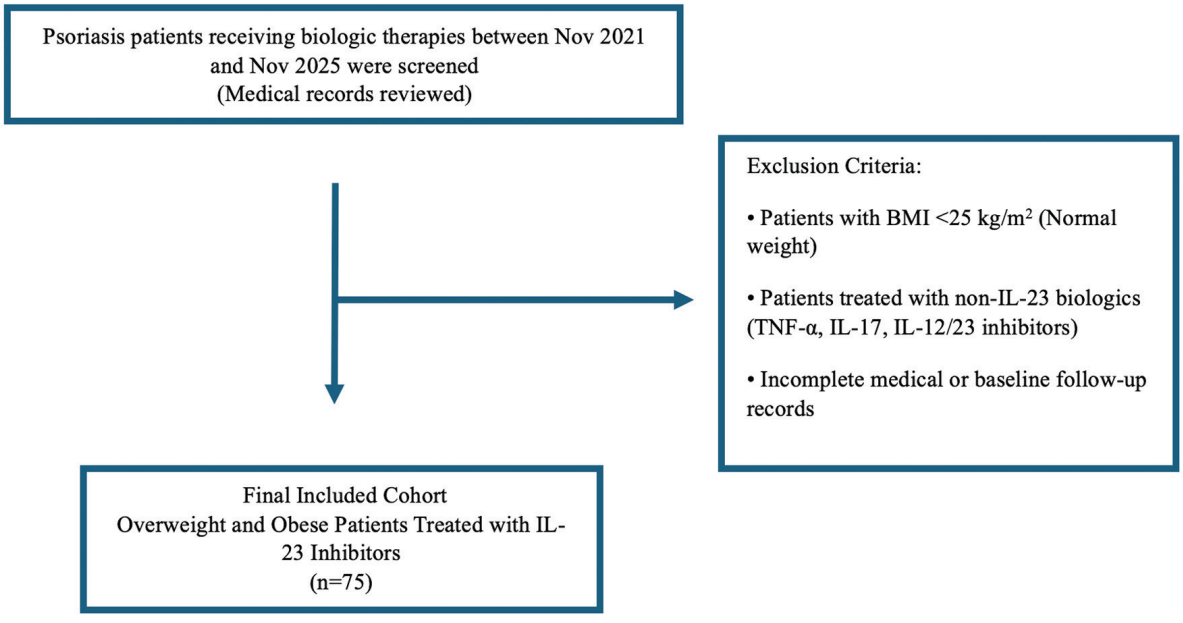

Figure 1. Flow diagram of patient selection and study cohort
BMI: Body mass index, IL: Interleukin, TNF-α: Tumor necrosis factor-α

Baseline demographic and clinical characteristics were collected, including age, sex, age at disease onset, family history of psoriasis, body weight, and BMI. In addition, data on psoriatic arthritis, involvement of special sites (scalp, palmoplantar, inverse, and nail psoriasis), and presence of systemic comorbidities were recorded.

Treatment-related data included prior conventional systemic therapies, prior biologic use, and prior phototherapy. Disease severity and therapeutic response were evaluated using the Psoriasis Area and Severity Index (PASI) at baseline and at weeks 4, 12-16, 28, and 52. Treatment responses were categorized as PASI75, PASI90, and PASI100, corresponding to ≥75%, ≥90%, and 100% improvement from baseline PASI scores, respectively.

Safety was assessed based on adverse events reported during routine clinical visits, physical examination findings, and laboratory monitoring. Hematological and biochemical analyses were performed every 3 months. Serology for hepatitis B, hepatitis C, and HIV was assessed at baseline and, in patients with negative baseline results, annually.

Patients with serological evidence of prior hepatitis B exposure [hepatitis B core antibody (anti-HBc) immunoglobulin G (IgG) positivity] underwent hepatitis B virus (HBV) deoxyribonucleic acid testing and were classified accordingly. These patients were co-managed by infectious disease or gastroenterology specialists, and laboratory monitoring was performed every 3 months.

Interferon gamma release assay (QuantiFERON-TB Gold) testing was performed at baseline and annually in patients with initially negative results. Chest radiography was performed at baseline and repeated at 6-month intervals during follow-up. Patients with positive QuantiFERON results were evaluated in collaboration with pulmonology specialists to rule out active tuberculosis and determine the need for latent tuberculosis infection (LTBI) prophylaxis.

For patients who discontinued treatment within the 52-week follow-up period, events leading to discontinuation were recorded. Primary cutaneous non-response was defined as failure to achieve a PASI75 response at weeks 12-16. Secondary cutaneous non-response was defined as loss of an initially achieved PASI75 response during follow-up. In patients with psoriatic arthritis, failure to achieve or maintain adequate control of joint symptoms was also considered treatment failure and classified as either primary joint non-response (within the first 12-16 weeks) or secondary joint non-response (loss of initially achieved disease control).

Statistical Analysis

Statistical analyses were conducted using IBM SPSS Statistics for Windows, version 26.0 (IBM Corp., Armonk, NY, USA). The normality of continuous variables was assessed using visual methods (histograms and Q-Q plots) and the Shapiro-Wilk test. Normally distributed continuous variables were presented as mean ± standard deviation, whereas non-normally distributed variables were presented as median (interquartile range). Categorical variables were presented as number (n) and percentage (%). Comparisons between two independent groups were performed using the independent samples t-test for normally distributed continuous variables and the Mann-Whitney U test for non-normally distributed continuous variables. Categorical variables were compared using the Pearson chi-square test or Fisher's exact test, as appropriate.

Drug survival was evaluated using Kaplan-Meier survival analysis, and differences between groups were assessed using the log-rank test. Univariable logistic regression analyses were performed to evaluate potential predictors of achieving PASI90 at week 52. Variables associated with the outcome in univariable analysis, together with clinically relevant potential confounders, were entered into the multivariable logistic regression model. Model calibration was assessed using the Hosmer-Lemeshow goodness-of-fit test, and model discrimination was evaluated using the area under the receiver operating characteristic curve (ROC-AUC). Odds ratios (ORs) with 95% confidence intervals (CIs) were reported. A two-sided p-value <0.05 was considered statistically significant.

Results

A total of 75 overweight or obese patients with psoriasis who were treated with IL-23 inhibitors were included in the study; 38 (50.7%) received guselkumab, and 37 (49.3%) received risankizumab. The most commonly involved special areas were the scalp (93.3%) and nails (88.0%). The most frequent comorbidities were hypertension (36.0%), hepatosteatosis (33.3%), dyslipidemia (32.0%), diabetes mellitus (25.3%), and coronary artery disease (22.7%). A total of 42.7% of patients had previously been treated with biologics, most commonly with ustekinumab (20.0%), secukinumab (16.0%), and ixekizumab (13.3%). Baseline demographics and clinical features are summarized in Table 1.

The mean baseline PASI score was 13.7±7.7 (Table 2). At week 4, the mean PASI score decreased to 3.6±3.3 and to 2.2±3.0 at weeks 12-16. PASI90 and PASI100 response rates increased progressively from 37.3% and 12.0% at weeks 12-16 to 71.6% and 44.6% at week 28, reaching 77.5% and 56.3% at week 52, respectively. Detailed response rates at all time points are presented in Table 2. Exploratory subgroup analyses are summarized in Supplementary Tables S1 and S2. Overweight patients demonstrated significantly higher PASI90 response rates at weeks 28 and 52 than those of obese patients, whereas no statistically significant differences were observed in

Table 1. Baseline demographic and clinical characteristics of patients	
	Total group (n = 75)
Gender, n (%): male/female	46 (61.3)/29 (38.7)
Age, years, mean ± SD	52.3±14.2
Age at disease onset, years, mean ± SD	32.1±15.3
Psoriasis in first-degree relatives, n (%)	25 (33.3)
Body mass index, mean ± SD	29.9±4.0
Biologic-experienced patients, n (%)	32 (42.7)
Baseline PASI, mean ± SD	13.7±7.7
Total treatment duration, months, mean ± SD	34.6±11.6
Psoriatic arthritis, n (%)	38 (50.7)
Involvement of special areas	**n (%)**
Scalp	70 (93.3)
Nail	66 (88.0)
Inverse	10 (13.3)
Palmoplantar	7 (9.3)
Comorbidities	**n (%)**
Hypertension	27 (36.0)
Hepatosteatosis	25 (33.3)
Dyslipidemia	24 (32.0)
Diabetes mellitus	19 (25.3)
Coronary artery disease	17 (22.7)
Thyroid disease	6 (8.0)
Psychiatric comorbidities	5 (6.7)
Cerebrovascular disease	1 (1.3)
Inflammatory bowel disease	1 (1.3)
Previous cutaneous malignancy	1 (1.3)
Previous non-cutaneous malignancy	0 (0.0)
Latent tuberculosis infection n (%)	30 (40.0)
Hepatitis B status	**n (%)**
Isolated anti-HBc IgG positivity	3 (4.0)
Resolved hepatitis B infection	13 (17.3)
Non-immune (susceptible)	45 (60.0)
Vaccinated	14 (18.7)
Previous systemic treatments	**n (%)**
Narrowband UVB	31 (41.3)
Methotrexate	58 (77.3)
Acitretin	33 (44.0)
Cyclosporine	15 (20.0)
Etanercept	5 (6.7)
Adalimumab	7 (9.3)
Infliximab	2 (2.7)
Certolizumab	3 (4.0)
Ustekinumab	15 (20.0)
Secukinumab	12 (16.0)
Ixekizumab	10 (13.3)
Guselkumab	2 (2.7)
Risankizumab	3 (4.0)
BMI: Body mass index, NB-UVB: Narrowband ultraviolet B, PASI: Psoriasis Area and Severity Index, SD: Standard deviation, y: Years, IgG: Immunoglobulin G	

Table 2. Treatment outcomes during IL-23 inhibitor therapy					
IL-23 inhibitor treatment			**n (%)**		
Guselkumab			38 (50.7)		
Risankizumab			37 (49.3)		
Total treatment duration, months, mean ± SD			34.6±11.6		
Time point	**n**	**Absolute PASI (mean ± SD)**	**PASI75 n (%)**	**PASI90 n (%)**	**PASI100 n (%)**
Baseline	75	13.7±7.7	-	-	-
Week 4	75	3.6±3.3	46 (61.3)	5 (6.7)	2 (2.7)
Week 12-16	75	2.2±3.0	71 (94.7)	28 (37.3)	9 (12.0)
Week 28	74	1.1±1.5	73 (98.6)	53 (71.6)	33 (44.6)
Week 52	71	0.8±1.1	68 (95.8)	55 (77.5)	40 (56.3)

PASI75/90/100 indicate ≥75%, ≥90%, and 100% improvement from baseline PASI score
PASI: Psoriasis Area and Severity Index, SD: Standard deviation

PASI75 or PASI100 responses between overweight and obese patients (Supplementary Table S1). Similarly, risankizumab-treated patients achieved higher PASI response rates at several early and intermediate time points, whereas no statistically significant differences were observed at week 52 (Supplementary Table S2). Univariable logistic regression analyses are presented in Supplementary Table S3. Lower BMI and biologic-naïve status were significantly associated with achieving PASI90 at week 52, whereas age, sex, psoriatic arthritis, and comorbidities were not. In multivariable logistic regression analysis (Table 3), increasing BMI (adjusted OR 0.773, 95% CI 0.640-0.933; p=0.007) and previous biologic exposure (adjusted OR 0.149, 95% CI 0.037-0.606; p=0.008) were independently associated with lower odds of achieving PASI90 at week 52 after adjustment for diabetes mellitus and hepatic steatosis. The model demonstrated acceptable discrimination (ROC-AUC=0.789) and good calibration, as assessed by the Hosmer-Lemeshow goodness-of-fit test (χ^2=6.303, df=8, p=0.613).

Hepatitis B infection was present in 21.3% (n=16) of patients (Table 1). All three patients with isolated anti-HBc IgG positivity received antiviral prophylaxis. Among the 13 patients with resolved hepatitis B infection, 10 received prophylaxis. No evidence of HBV reactivation was detected in any patient over the course of follow-up.

Latent tuberculosis infection was detected in 40.0% (n=30) of patients. All patients started prophylaxis; it was successfully completed in all but one, who discontinued treatment due to adverse effects. No cases of LTBI reactivation were observed.

The most commonly observed adverse events were infections, including upper respiratory tract infections (6.7%) and dermatophyte or Candida infections (2.7%) (Table 4). One case was newly diagnosed with basal cell carcinoma during follow-up.

Treatment was discontinued in 4 patients (5.3%) during the 52-week follow-up period (Table 4), primarily due to secondary non-response (2.7%), followed by primary non-response (1.3%) and adverse events (1.3%). The overall 52-week drug survival rate was 94.7% (Figure 2). Kaplan-Meier analysis showed no significant difference in drug survival between overweight and obese patients (log-rank p=0.143) or between patients treated with guselkumab and those treated with risankizumab (log-rank p=0.319).

Discussion

In this real-world study, IL-23 inhibitors demonstrated high and sustained efficacy in psoriasis patients with higher BMI. At weeks 12-16, PASI90 and PASI100 responses were attained in 37.3% and 12.0% of patients, respectively; these increased to 77.5% and 56.3% at week 52. In phase 3 clinical trials, both guselkumab and risankizumab have demonstrated high efficacy, with substantial PASI90 and PASI100 response rates as early as week 16 (13,14). Although response rates at earlier time points in the present study were lower than those reported in clinical trials, treatment outcomes at week 52 (PASI90: 77.5%; PASI100: 56.3%) were comparable to the efficacy rates

Table 3. Multivariable logistic regression analysis of predictors of achieving PASI90 at week 52		
Variable	**Adjusted OR (95% CI)**	**p**
BMI (continuous, kg/m^2)	0.773 (0.640-0.933)	0.007
Previous biologic exposure	0.149 (0.037-0.606)	0.008
Diabetes mellitus	1.516 (0.308-7.471)	0.609
Hepatosteatosis	0.610 (0.151-2.464)	0.487

Variables entered into the model: BMI (continuous, kg/m^2), previous biologic exposure, diabetes mellitus, and hepatic steatosis. The model demonstrated acceptable discrimination (area under the receiver operating characteristic curve [AUC]=0.789) and good calibration according to the Hosmer-Lemeshow goodness-of-fit test (χ^2=6.303, df=8, p=0.613).
BMI: Body mass index, CI: Confidence interval, OR: Odds ratio, PASI: Psoriasis Area and Severity Index

Table 4. Adverse events and reasons for treatment discontinuation	
Adverse events	**n (%)**
Upper respiratory tract infection	5 (6.7)
Urinary tract infection	1 (1.3)
Dermatophyte or Candida infection	2 (2.7)
Non-melanoma skin cancer	1 (1.3)
Hepatitis B reactivation n (%)	**0 (0.0)**
Tuberculosis reactivation n (%)	**0 (0.0)**
Reasons for treatment discontinuation	**n (%)**
Primary cutaneous non-response	1 (1.3)
Secondary cutaneous non-response	1 (1.3)
Secondary joint non-response	1 (1.3)
Non-melanoma skin cancer (basal cell carcinoma)	1 (1.3)

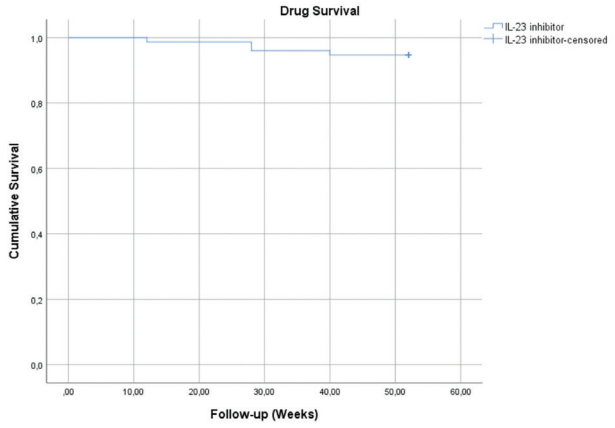

Figure 2. Kaplan-Meier curve showing the 52-week drug survival of the study cohort
IL: Interleukin

observed in clinical studies of both agents (13,14). It should be noted, however, that the mean baseline PASI score in our cohort (13.7) was considerably lower than those reported in the phase 3 trials of risankizumab (20.6) and guselkumab (21.9) (13,14). This likely reflects routine clinical practice in which biologic therapy may be initiated for patients with lower PASI scores owing to involvement of special areas, psoriatic arthritis, impaired quality of life, or previous treatment failure. In line with this, scalp and nail involvement were highly prevalent in our cohort, affecting 93.3% and 88.0% of patients, respectively. In addition, the lower baseline PASI may have partly contributed to the high PASI100 rates observed at week 52.

Pooled analyses of phase 3 studies of IL-23 inhibitors have shown that treatment responses are generally consistent across BMI categories, although slightly reduced responses have been reported in obese patients compared to individuals with lower BMI (13,15). These findings indicate that IL-23 inhibitors remain highly effective despite the potential adverse effects of obesity on treatment outcomes.

Existing real-world evidence regarding the influence of BMI on the effectiveness of IL-23 antagonists has yielded conflicting results. In a study assessing predictors of achieving super-responder status (complete clearance at week 20) with guselkumab, obesity was identified as a negative predictor, along with prior biologic experience and higher baseline PASI scores (16). Similarly, a real-world registry study demonstrated that although patients across all BMI categories experienced substantial improvements with guselkumab, obese patients were less likely to achieve relative endpoints such as PASI90 or complete clearance compared to those with lower BMI (17). Another study demonstrated reduced complete clearance rates at week 16 in obese patients compared to individuals with lower BMI; however, this difference attenuated over time, suggesting that obesity may be associated with a delayed response to risankizumab (18). Similarly, heavier patients have been shown to be less likely to achieve a PASI75 response at early time points, suggesting a delayed response to guselkumab.

In contrast, other real-world studies evaluating IL-23 inhibitors, including guselkumab, risankizumab, and tildrakizumab, have shown that overweight and obese patients can achieve comparable overall treatment responses, with mean PASI scores decreasing to low levels over time, suggesting that increased body weight may not significantly impair long-term efficacy (19-21).

Taken together, studies suggest that although obesity may be associated with a slower or less pronounced early response and a lower likelihood of achieving higher levels of skin clearance, IL-23 inhibitors appear to maintain robust and sustained efficacy in overweight and obese patients. Consistent with these observations, overweight patients in our cohort achieved significantly higher PASI90 response rates at weeks 28 and 52 than obese patients, whereas PASI75, PASI100, and drug survival were comparable between the two BMI groups. Furthermore, lower BMI remained independently associated with achieving PASI90 at week 52 after adjustment for previous biologic exposure, diabetes mellitus, and hepatosteatosis. This finding suggests that increasing BMI may independently reduce the likelihood of achieving PASI90, even after accounting for potential metabolic confounders such as diabetes mellitus and hepatosteatosis. Similarly, risankizumab-treated patients achieved higher PASI response rates at several early and intermediate time points than did guselkumab-treated patients, whereas no significant differences were observed between the groups at week 52 or in drug survival. These findings may suggest differences in the timing of clinical response

rather than long-term effectiveness; however, as this was an exploratory comparison without adjustment for potential baseline differences between treatment groups, no definitive conclusions can be drawn. Larger multicenter comparative studies are warranted to determine whether clinically meaningful differences exist between individual IL-23 inhibitors in overweight and obese patients.

In addition, IL-23 inhibitors demonstrated a favorable safety profile, consistent with previous reports (22-24). Infections were the most commonly observed adverse events. No adverse events led to treatment termination; the only exception was one case of newly diagnosed basal cell carcinoma. A considerable proportion of patients had concomitant hepatitis B infection (n=16, 21.3%). Prophylaxis rates were high, with 13 of 16 patients receiving antiviral therapy. None of the cases developed reactivation of the HBV during follow-up, supporting the safety of IL-23 inhibitors. These results align with earlier research indicating that the risk of HBV reactivation remains minimal during treatment with IL-23 inhibitors (25-27). Similarly, LTBI was common (n=30, 40.0%). Prophylaxis was initiated in all patients and completed in all but one due to adverse effects. None of the cases showed reactivation. In accordance with available clinical research and observational real-world data, these results highlight the favorable safety profile of IL-23 inhibitors regarding their low potential for reactivation (28-32).

Study Limitations

This was a single-center study with a relatively small sample size and a retrospective design, which constituted the main limitations. In addition, the absence of a normal-weight comparator group precludes direct comparisons across BMI categories and limits conclusions regarding the independent effect of BMI on treatment outcomes. However, the study provides valuable real-world clinical data on the effectiveness, safety, and drug survival of IL-23 inhibitors in an overweight and obese population with psoriasis, a group known to have a reduced response to treatment. Larger, multicenter comparative studies including patients from all BMI categories are warranted. Additionally, the small number of treatment discontinuations during follow-up limited the statistical power of the drug survival analyses and subgroup comparisons; therefore, these findings should be interpreted with caution. Despite these limitations, the present study provides valuable real-world evidence supporting the sustained effectiveness, favorable safety profile, and high drug survival of IL-23 inhibitors in overweight and obese patients with psoriasis.

Conclusion

IL-23 inhibitors demonstrated sustained clinical effectiveness, high 52-week drug survival, and a favorable safety profile in this real-world cohort of overweight and obese patients with psoriasis. Although increasing BMI was independently associated with a lower likelihood of achieving PASI90 at week 52, drug survival remained comparable across BMI groups. These findings support IL-23 inhibitors as an effective, well-tolerated treatment option for overweight and obese patients with psoriasis in routine clinical practice.

Ethics

Ethics Committee Approval: The study received approval from the University of Health Sciences Türkiye, Haseki Training and Research Hospital Non-Interventional Clinical Research Ethics Committee (approval number: 106-2026, date: 22.04.2026) and was conducted in accordance with the Declaration of Helsinki.

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

Authorship Contributions

Surgical and Medical Practices: G.P.U., Y.S., Concept: G.P.U., Design: G.P.U., Data Collection or Processing: G.P.U., Y.S., Analysis or Interpretation: G.P.U., Y.S., Literature Search: G.P.U., Y.S., Writing: G.P.U., Y.S.

Conflict of Interest: No conflicts of interest were declared by the authors.

Financial Disclosure: This study received no financial support.

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

Supplementary Tables Link: https://d2v96fxpocvxx.cloudfront.net/7a593d95-86ec-4d2b-8dd3-26b0d0b79ea4/content-images/67b21eb4-07d0-4ea2-9ddc-76b4d2beb3cb.pdf

References

1. Parisi R, Iskandar IYK, Kontopantelis E, et al. National, regional, and worldwide epidemiology of psoriasis: systematic analysis and modelling study. BMJ. 2020;369:m1590.
2. Yaylı S, Topbaş M, Aksu Arıca D, Tuğcugil S, Çapkın E, Bahadır S. Trabzon ilinde psoriasis prevalansı. Turkderm. 2016;50:141-4.
3. Karmbir, Faisal SM, Narang RK, Kurmi BD, Kaur K, Sharma A. Psoriasis beyond the skin: systemic inflammation as a bridge to metabolic and hepatic comorbidities. Inflammopharmacology. 2026;34:1317-30.
4. Ma J, Li H, Yang X, et al. Psoriasis as a systemic inflammatory disease: an immune set-point framework for comorbidities and relapse. Front Immunol. 2026;17:1830099.

5. Armstrong AW, Harskamp CT, Armstrong EJ. Psoriasis and the risk of diabetes mellitus: a systematic review and meta-analysis. JAMA Dermatol. 2013;149:84-91.
6. Li W, Han J, Qureshi AA. Obesity and risk of incident psoriatic arthritis in US women. Ann Rheum Dis. 2012;71:1267-72.
7. Sbidian E, Chaimani A, Guelimi R, et al. Systemic pharmacological treatments for chronic plaque psoriasis: a network meta-analysis. Cochrane Database Syst Rev. 2025;8:CD011535.
8. Burshtein J, Armstrong A, Chow M, et al. The association between obesity and efficacy of psoriasis therapies: an expert consensus panel. J Am Acad Dermatol. 2025;92:807-15.
9. Prussick R, Unnebrink K, Valdecantos WC. Efficacy of adalimumab compared with methotrexate or placebo stratified by baseline BMI in a randomized placebo-controlled trial in patients with psoriasis. J Drugs Dermatol. 2015;14:864-8.
10. Giunta A, Babino G, Ruzzetti M, Manetta S, Chimenti S, Esposito M. Influence of body mass index and weight on etanercept efficacy in patients with psoriasis: a retrospective study. J Int Med Res. 2016;44:72-5.
11. Enos CW, Ramos VL, McLean RR, et al. Comorbid obesity and history of diabetes are independently associated with poorer treatment response to biologics at 6 months: a prospective analysis in Corrona Psoriasis Registry. J Am Acad Dermatol. 2022;86:68-76.
12. Zweegers J, van den Reek JM, van de Kerkhof PC, et al. Body mass index predicts discontinuation due to ineffectiveness and female sex predicts discontinuation due to side-effects in patients with psoriasis treated with adalimumab, etanercept or ustekinumab in daily practice: a prospective, comparative, long-term drug-survival study from the BioCAPTURE registry. Br J Dermatol. 2016;175:340-7.
13. Strober B, Menter A, Leonardi C, et al. Efficacy of risankizumab in patients with moderate-to-severe plaque psoriasis by baseline demographics, disease characteristics and prior biologic therapy: an integrated analysis of the phase III UltIMMa-1 and UltIMMa-2 studies. J Eur Acad Dermatol Venereol. 2020;34:2830-8.
14. Reich K, Armstrong AW, Foley P, et al. Efficacy and safety of guselkumab, an anti-interleukin-23 monoclonal antibody, compared with adalimumab for the treatment of patients with moderate to severe psoriasis with randomized withdrawal and retreatment: results from the phase III, double-blind, placebo- and active comparator-controlled VOYAGE 2 trial. J Am Acad Dermatol. 2017;76:418-31.
15. Gordon KB, Blauvelt A, Foley P, et al. Efficacy of guselkumab in subpopulations of patients with moderate-to-severe plaque psoriasis: a pooled analysis of the phase III VOYAGE 1 and VOYAGE 2 studies. Br J Dermatol. 2018;178:132-9.
16. Mortato E, Talamonti M, Marcelli L, et al. Predictive factors for super responder status and long-term effectiveness of guselkumab in psoriasis: a multicenter retrospective study. Dermatol Ther(Heidelb). 2025;15:1239-50.
17. Armstrong AW, Fitzgerald T, McLean RR, et al. The effectiveness of guselkumab by BMI category among patients with moderate-to-severe plaque psoriasis in the corevitas psoriasis registry. Adv Ther. 2023;40:2493-508.
18. Dragotto M, Capalbo E, Cartocci A, et al. Real-life effectiveness of risankizumab according to body mass index: results of an Italian multicentre retrospective study. J Eur Acad Dermatol Venereol. 2023.
19. Ricceri F, Chiricozzi A, Peris K, Prignano F. Successful use of anti-IL-23 molecules in overweight-to-obese psoriatic patients: a multicentric retrospective study. Dermatol Ther. 2022;35:e15793.
20. Rompoti N, Stefanaki I, Panagakis P, et al. Real-world effectiveness, drug survival and safety of risankizumab over a period of 2 years in 158 patients with moderate-to-severe plaque psoriasis. J Eur Acad Dermatol Venereol. 2025;39:e1-4.
21. Gargiulo L, Ibba L, Pavia G, et al. Real-life effectiveness and safety of risankizumab in 131 patients affected by moderate-to-severe plaque psoriasis: a 52-week retrospective study. Dermatol Ther(Heidelb). 2022;12:2309-24.
22. Olivares-Guerrero M, González-Quesada A, Sahuquillo-Torralba A, et al. Risk of infections in patients with psoriasis treated with biologic agents and new oral small molecules: BIOBADADERM registry. J Invest Dermatol. 2026.
23. Mastorino L, Dapavo P, Ortoncelli M, et al. Drug survival, effectiveness and safety of risankizumab for moderate-to-severe psoriasis for up to 4 years. Clin Exp Dermatol. 2025;51:35-41.
24. Atzeni F, Rotondo C, Siragusano C, et al. Italian multicenter real-world study on the twelve-month effectiveness, safety, and retention rate of guselkumab in psoriatic arthritis patients. J Clin Med. 2025;14:4111.
25. Chiu HY, Hsu CC, Tsai TF, et al. Comparative risk of reactivation of hepatitis B and C after treatment with biologics and targeted synthetic DMARDs in psoriasis and psoriatic arthritis: a 15-year multicenter cohort study. J Am Acad Dermatol. 2026;94:458-66.
26. Ciolfi C, Balestri R, Bardazzi F, et al. Safety profile of risankizumab in the treatment of psoriasis patients with concomitant hepatitis B or C infection: a multicentric retrospective cohort study of 49 patients. J Eur Acad Dermatol Venereol. 2023;37:e1203-7.
27. Huang YH, Yen JS, Li SH, Chiu HY. Safety profile of guselkumab in treatment of patients with psoriasis and coexisting hepatitis B or C: a multicenter prospective cohort study. J Am Acad Dermatol. 2024;90:1083-6.
28. Raimondo A, Amerio P, Balato A, et al. Risankizumab efficacy and safety in psoriatic patients with latent tuberculosis infection: a multicentric real-world study. J Eur Acad Dermatol Venereol. 2025;39:e287-9.
29. Yuan L, Li Y, Lin J, et al. Safety of biologics for psoriatic patients with latent tuberculosis. Clin Cosmet Investig Dermatol. 2023;16:2333-6.

30. Mortato E, Talamonti M, Marcelli L, et al. A long-term real-world safety study of guselkumab in patients with psoriasis who have infectious comorbidities, malignancies or heart disease: the EARLY study. Clin Exp Dermatol. 2025;50:1786-94.

31. Torres T, Chiricozzi A, Puig L, et al. Treatment of psoriasis patients with latent tuberculosis using IL-17 and IL-23 inhibitors: a retrospective, multinational, multicentre study. Am J Clin Dermatol. 2024;25:333-42.

32. Nogueira M, Warren RB, Torres T. Risk of tuberculosis reactivation with interleukin(IL)-17 and IL-23 inhibitors in psoriasis - time for a paradigm change. J Eur Acad Dermatol Venereol. 2021;35:824-34.

Original Article

DOI: 10.4274/haseki.galenos.2026.20592
Med Bull Haseki 2026;64(4):354-361

Serum PADI4 is Elevated in Neuro-Behçet's Disease but is not Associated with Clinical or Cognitive Outcomes

Cem Ismail Kucukali[1], Nuray Sancar[1], Melike Tezel[1], Hande Yuceer Korkmaz[1], Tuncay Gunduz[2], Murat Kurtuncu[2]

[1]*Istanbul University Aziz Sancar Institute of Experimental Medicine, Department of Neuroscience, Istanbul, Türkiye*
[2]*Istanbul University, Istanbul Faculty of Medicine, Department of Neurology, Istanbul, Türkiye*

Abstract

Aim: Peptidylarginine deiminase type 4 (PADI4), a key enzyme involved in neutrophil extracellular trap (NET) formation, has been implicated in the pathogenesis of autoinflammatory diseases, including Behçet's disease (BD). We aimed to evaluate serum PADI4 levels in neuro-BD (NBD) and to investigate their association with disease activity and clinical parameters.

Methods: This study was designed as an analytical cross-sectional study between 2021 and 2023. A total of 25 patients with NBD, 45 with multiple sclerosis, 10 with neuromyelitis optica (NMO), and 19 healthy controls were enrolled. Serum PADI4 concentrations were measured using ELISA. Comprehensive neuropsychological assessments were conducted in the NBD group. Correlation analyses were performed to evaluate associations with cognitive and clinical parameters.

Results: Serum PADI4 levels were significantly elevated in NBD patients compared to healthy controls but were comparable among patient groups. None of the cognitive and motor performance scores were correlated with PADI4 levels. NBD patients with a positive pathergy test showed significantly higher PADI4 levels.

Conclusion: Neutrophil extracellular traposis may represent a shared inflammatory pathway across immune-mediated brain disorders. However, the lack of association between serum PADI4 levels and clinical or cognitive parameters in NBD suggests limited diagnostic and prognostic utility as a biomarker.

Keywords: Peptidylarginine deiminase type 4, neutrophil extracellular trap, autoinflammatory disease, biomarker, Behçet's disease

Introduction

Neuro-Behçet's disease (NBD), an uncommon manifestation of Behçet's disease (BD), is an inflammatory disorder characterized by recurrent and unpredictable relapsing-remitting episodes of neurological involvement occurring in approximately 5-10% of individuals diagnosed with BD (1). According to magnetic resonance imaging (MRI) findings, NBD can be classified as parenchymal or non-parenchymal and may lead to neurocognitive impairments, particularly in attention, delayed recall, working memory, and executive functioning (2,3). BD is considered to have predominantly autoinflammatory features, involving dysregulated neutrophil activity (4). Neutrophils contribute to inflammation by generating neutrophil extracellular traps (NETs), which are released through a mechanism called NETosis and accumulate in serum and inflamed tissues, thereby promoting endothelial damage and a procoagulant state (5). The pathergy test, a diagnostic tool for BD, is characterized by non-specific skin hypersensitivity and intensive neutrophil infiltration, which clinically mirrors this hyperinflammatory profile (6).

Peptidylarginine deiminase type 4 (PADI4), a calcium ion-dependent enzyme, catalyzes protein citrullination by converting arginine to citrulline, thereby altering protein structure and function (7). It plays a critical role in inflammatory processes by mediating histone citrullination

Corresponding Author: Prof. Dr. Cem Ismail Kucukali, Istanbul University Aziz Sancar Institute of Experimental Medicine, Department of Neuroscience, Istanbul, Türkiye
E-mail: cemsmile@gmail.com **ORCID:** orcid.org/0000-0001-9851-8577
Received: 22.04.2025 **Accepted:** 10.03.2026 **Publication Date:** 24.09.2026

Cite this article as: Kucukali CI, Sancar N, Tezel M, Yuceer Korkmaz H, Gunduz T, Kurtuncu M. Serum PADI4 is elevated in Neuro-Behçet's disease but is not associated with clinical or cognitive outcomes. Med Bull Haseki. 2026;64(4):354-361

during NET formation and has been associated with many autoimmune conditions due to its effects on immune cell dysregulation (8). In multiple sclerosis (MS), PADI4 significantly facilitates the citrullination of myelin basic protein, leading to immune-mediated tissue damage. The direct association of PADI4 with BD remains unconfirmed; however, its role in critical neuro-inflammatory and autoimmune pathways emphasizes its potential relevance in the disease's pathophysiology (9). Therefore, we hypothesized that PADI4-mediated citrullination and NETosis lead to neuroinflammatory damage in NBD, and that PADI4 levels might correlate with disease activity.

This study aimed to investigate the role of PADI4 in NBD, with a particular focus on its involvement in immune-inflammatory processes and its potential association with disease activity. Serum PADI4 levels were compared among patients with NBD, other inflammatory brain disorders, and healthy controls, and their relationships with clinical and cognitive parameters were evaluated.

Materials and Methods

Compliance with Ethical Standards

Ethical approval was taken from the Istanbul University Faculty of Medicine Clinical Research Ethics Committee (approval no: 09, date: 16.04.2021). The study protocol was approved by the Institutional Review Board. The study was conducted in accordance with the ethical principles of the Declaration of Helsinki. All subjects provided written informed consent prior to any study-related procedure.

Study Design

This study was designed as an analytical cross-sectional study between 2021 and 2023. A total of 25 patients diagnosed with NBD according to the International Criteria for BD were enrolled in the study [median: 7.30 (interquartile ranges (IQR): 2.13); 14 men]. Additionally, 45 age- and gender-matched patients with [relapsing-remitting MS (RRMS), median: 6.48 (IQR: 2.92), 21 men], diagnosed according to the revised McDonald criteria, 10 patients with [neuromyelitis optica (NMO), median: 6.35 (IQR: 2.02), 5 men], and 19 healthy controls [healthy control, median: 3.58 (IQR: 1.97), 10 men] were included as control groups. Due to the exploratory nature of the study, a formal sample size calculation was not performed; however, all eligible patients during the study period were included (Figure 1).

Serum samples were obtained from all NBD patients and controls during remission and stored at -80 °C until use. Neuropsychological tests were administered during serum sampling. Magnetic resonance imaging and magnetic resonance venography confirmed typical parenchymal NBD findings—such as unilateral or bilateral hyperintense lesions extending from the brainstem to the diencephalon and basal ganglia—and showed no cerebral venous thrombosis indicative of non-parenchymal involvement.

ELISA

Human PADI4 levels were measured in serum samples using an ELISA kit (Bioassay Technology Laboratory, China) according to the manufacturer's instructions. Optical

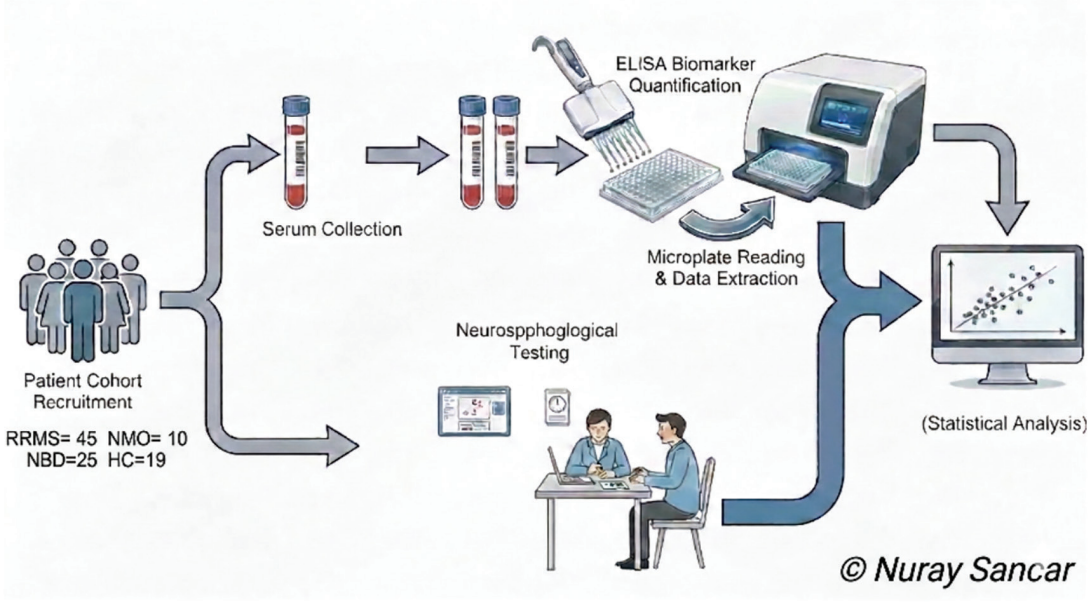

Figure 1. Flowchart of the study
RRMS: Relapsing-remitting multiple sclerosis, NMO: Neuromyelitis optica, NBD: Neuro-Behçet's disease, HC: Healthy controls, ELISA: Enzyme-linked immunosorbent assay

densities were measured at 450 nm, and concentrations were calculated based on standard curves provided in the kit. The results were expressed as ng/mL.

Neuropsychological Evaluation

General Cognitive Performance: The Mini-Mental State Examination and the Addenbrooke's Cognitive Examination-Revised (ACE-R) were administered to evaluate overall cognitive capacity.

Attention and Orientation: The ACE-R Attention and Orientation subtest and the Wechsler Memory Scale-Revised (WMS-R) Digit Span (forward and backward) were used to assess attention and working memory performance.

Memory The ACE-R Memory subtest, WMS-R Visual Memory (immediate recall, long-term recall, and recognition), and the California Verbal Learning Test-II were used to evaluate verbal and visual memory performance.

Language skills were assessed using the ACE-R Language subtest and the Boston Naming Test (spontaneous, semantic cue, phonemic cue, paraphasia, and functional knowledge).

The Clock Drawing Test, Stroop Test (interference times, errors, and spontaneous corrections), Trail Making Test (TMT) A and B (completion time and errors), Wisconsin Card Sorting Test, the Symbol Digit Modalities Test, and the Luria Perseveration Test were administered to assess executive functions.

The TMT A and B, the 9-Hole Peg Test (placement and removal times for dominant and non-dominant hands), and the Timed 25-Foot Walk Test were administered to evaluate psychomotor speed and mental flexibility.

Verbal and categorical fluency were assessed using the ACE-R Fluency subtest, the (Controlled Oral Word Association Test; animal category, perseverations, and out-of-category words), the (Category Fluency Score; total score, perseverations, out-of-category words, and proper names), and the KAS Total Score.

Visuospatial abilities were examined using the ACE-R Visuospatial subtest, the Benton Facial Recognition Test, the Benton Line Orientation Test, and the Figure Copying Test.

Statistical Analysis

Statistical analyses were performed using GraphPad Prism software. Data distribution was assessed using the Kolmogorov-Smirnov test and found to be non-normal; therefore, results are presented as median (interquartile range). Comparisons of serum PADI4 levels among NBD, RRMS, NMO, and healthy control groups were conducted using the Kruskal-Wallis test, followed by Dunn-Bonferroni post-hoc analysis. Pairwise comparisons were performed using the Mann-Whitney U test. Correlations between PADI4 levels and demographic, clinical, and neuropsychological parameters were evaluated using Spearman's rank correlation coefficient. A two-tailed p-value <0.05 was considered statistically significant, and Bonferroni correction was applied for multiple comparisons.

Results

Demographic and Clinical Characteristics of the NBD Group

The NBD group consisted of 25 patients (14 men and 11 women). The median duration of BD was 13 (IQR: 15), while the median duration of neurological involvement was 13 (IQR: 15). Pathergy testing was positive in 6 patients and negative in 13. Brain MRI findings revealed brainstem lesions in 15 patients, diencephalic involvement in 9 patients, and spinal cord lesions in 1 patient. The Expanded Disability Status Scale (EDSS) score had a median of 2 (IQR: 1.75). Detailed clinical and radiological characteristics of the NBD patients are presented in Table 1.

Table 1. Clinical, demographic features of NBD patients and controls

Feature	RRMS	NMO	NBD	HC
Age (years)	**Median: 52.0 (IQR: 9.5)**	**Median: 49.5 (IQR: 11.0)**	**Median: 52.0 (IQR: 4.0)**	**Median: 51.50 (IQR: 9.75)**
Gender (men/women)	21/24	5/5	14/11	10/9
BD duration (years)	NA	NA	Median: 22 (IQR:10)	NA
NBD duration (years)	NA	NA	Median: 13 (IQR:15)	NA
Pathergy test (positive/negative)*	NA	NA	6/13	NA
MRI lesions (brainstem)	ND	ND	15 patients	NA
MRI lesions (diencephalon)	ND	ND	9 patients	NA
MRI lesions (spinal cord)	ND	ND	1 patient	NA
EDSS	ND	ND	Median: 2 (IQR:1.75)	NA

*Note that the pathergy test was available for 19 patients. Continuous variables are presented as median (interquartile range).
RRMS: Relapsing remitting multiple sclerosis, NMO: Neuromyelitis optica, HC: Healthy control, ND: Not determined, NA: Not applicable, IQR: Interquartile range

Comparison of Serum PADI4 Levels Among NBD, RRMS, NMO, and Healthy Controls

The Kruskal-Wallis test yielded a significant p-value for multiple group comparisons of all groups (p<0.0001). Dunn's post-hoc test for multiple comparisons showed that NBD patients had significantly higher PADI4 levels than healthy controls (p<0.0001). Furthermore, both RRMS (p<0.0001) and NMO (p<0.0001) groups had significantly elevated levels of PADI4 compared with healthy controls. There were no differences among NBD, RRMS, and NMO groups in serum PADI4 levels. The graphical representation of these comparisons was provided in Figure 2.

Correlation Between PADI4 Levels and Clinical/ Cognitive Features in NBD

No significant correlations were found between PADI4 levels and age, disease duration, and EDSS score (p>0.05 for all).

To explore the potential association between serum PADI4 levels and cognitive performance in patients with NBD, a series of neuropsychological and motor performance tests was administered. These tests assessed a wide range of cognitive domains, including general cognitive status, memory, language, executive function, attention, visuospatial skills, and psychomotor speed. Median scores, IQR, and correlation p-values between PADI4 levels and individual test results were presented in Table 2.

Overall, correlation analyses revealed no statistically significant associations between PADI4 levels and neuropsychological test scores. However, weak but statistically significant correlations were observed between PADI4 levels and the Boston Naming Test–Semantic Cue condition (p=0.023) and the 9-Hole Peg Test (Non-Dominant Hand–Removal Time) (p=0.046). However, these p-values did not remain significant after Bonferroni correction, which required p-values less than 0.00001.

Evaluation of Pathergy Test Results and Association with PADI4 Levels in NBD

To assess the relationship between inflammatory response markers and immune hyperreactivity, PADI4 levels were compared between pathergy-positive and pathergy-negative NBD patients. Using the Mann-Whitney U test, a statistically significant difference in serum PADI4 levels was found between the two groups (p=0.046, two-tailed, exact p-value). Patients who were pathergy-positive exhibited higher median PADI4 levels (median =8.434 ng/mL, n=6) compared to those who were pathergy-negative (median =6.91 ng/mL, n=13). The graphical representation of this comparison is shown in Figure 3.

Discussion

To our knowledge, this study provides additional evidence that serum PADI4 levels are elevated in patients with NBD compared to healthy controls. Furthermore, higher PADI4 levels were observed in patients with positive pathergy tests. Thus, our study provides new evidence supporting the role of PADI4 in the immunopathology of NBD. On the other hand, our study and previous studies have also shown that PADI4 elevation is not specific to NBD or BD and may also be observed in other immune-mediated disorders such as systemic lupus erythematosus, rheumatoid arthritis, MS and NMO (5,9-11). These findings

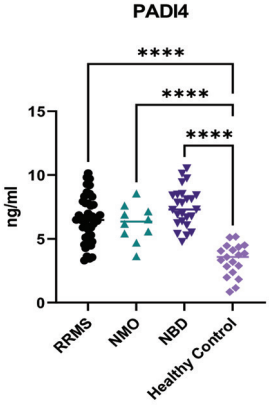

Figure 2. Comparison of serum PADI4 levels across study groups. Horizontal lines indicate median values. The Kruskal-Wallis test yielded a significant p-value for the comparison among all groups (p<0.0001). Asterisks denote significant p-values by Dunn's post-hoc test, ****p<0.0001

RRMS: Relapsing-remitting multiple sclerosis, NMO: Neuromyelitis optica, NBD: Neuro-Behçet's disease, PADI4: Peptidylarginine deiminase type 4

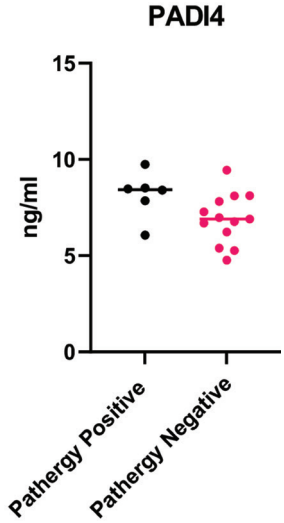

Figure 3. Comparison of serum PADI4 levels between pathergy-positive and pathergy-negative NBD patients.
The horizontal line indicates median values. The p-value (p=0.046) was obtained by the Mann-Whitney U test

PADI4: Peptidylarginine deiminase type 4, NBD: Neuro-Behçet's disease

Table 2. Neuropsychological test scores of NBD patients: median, interquartile range (IQR), p and r values of correlation tests with PADI4 levels

Test name	Median	IQR	p-value	r-value
MMSE/max 30	25.0	7.0	0.517	0.142
ACE-R total/max 100	69.0	19.0	0.442	0.168
ACE-R attention and orientation/max 18	16.0	6.0	0.478	0.155
ACE-R memory/max 26	12.0	7.0	0.662	0.096
ACE-R fluency/max 14	10.0	4.0	0.446	0.167
ACE-R language/max 26	20.0	9.0	0.314	0.207
ACE-R visuospatial/max 16	3.0	3.0	0.297	0.226
WMS-R digit span (forward)	5.0	0.5	0.473	0.186
WMS-R digit span (backward)	4.0	1.0	0.966	0.010
WMS-R visual memory (immediate)	7.0	6.0	0.381	0.226
WMS-R visual memory (long-term recall)	5.0	4.5	0.950	0.016
WMS-R visual memory (long-term recognition)	2.0	1.5	0.468	0.188
Verbal fluency (fruit-name)	6.5	5.0	0.375	0.229
COWAT animal score	16.0	7.5	0.268	0.284
COWAT animal perseveration	0.6	1.0	0.665	-0.113
KAS total score	19.0	23.0	0.740	-0.086
KAS perseveration	1.0	1.5	0.456	-0.193
KAS out of category	0.1	0.3	0.556	-0.153
KAS proper name	0.8	1.0	0.735	0.088
Stroop interference time	66.5	41.5	0.853	-0.054
Stroop spontaneous correction	2.5	3.25	0.559	-0.170
Stroop error	3.50	9.25	0.451	0.219
Trail making test A - time	82.0	42.75	0.969	0.010
Trail making test A - errors	0.2	0.4	0.542	-0.164
Trail making test B - time	255	132.5	0.716	-0.111
Trail making test B - errors	2.0	2.0	0.160	0.413
Clock drawing test	5.0	2.0	0.143	0.370
Luria perseveration (present:1 absent: 0)	0.4	1.0	0.756	-0.081
Symbol digit modalities test (SDMT) total items	21.0	14.5	0.202	0.336
Symbol digit modalities test total correct	20.5	16.5	0.261	0.298
Figure copying test	3.0	2.0	0.444	0.198
CVLT II total correct	33.0	16.5	0.980	-0.006
CVLT II B list	4.0	2.5	0.552	-0.155
CVLT II short-term cued recall	6.0	5.0	0.450	-0.196
CVLT II long-term cued recall	7.0	4.5	0.951	0.016
CVLT II short-term uncued recall	4.0	5.5	0.711	0.096
CVLT II long-term uncued recall	7.0	5.0	0.358	-0.237
CVLT II total	9.0	12.0	0.453	-0.195
CVLT II recognition	13.0	6.5	0.422	0.208
CVLT II recognition (YP)	4.0	5.5	0.346	-0.243
Boston naming test - spontaneous	23.0	8.0	0.323	0.255
Boston naming test - no naming	1.0	7.5	0.322	-0.255
Boston naming test - semantic cue	2.0	2.0	**0.023**	-0.544
Boston naming test - phonemic cue	2.0	3.0	0.596	0.138

Table 2. Continued				
Test name	**Median**	**IQR**	**p-value**	**r-value**
Boston naming test - function	0.5	0.9	0.105	0.406
Boston naming test - paraphasia	0.0	0.5	0.215	0.316
Benton facial recognition test	43.0	9.0	0.437	0.201
Benton line orientation test	17.5	6.5	0.458	0.199
Wisconsin card sorting test (WCST) total responses	125.5	34	0.981	0.006
Wisconsin card sorting test total errors	40.0	35.25	0.850	-0.051
Wisconsin card sorting test total correct	74.5	27.0	0.832	0.057
Wisconsin card sorting test categories completed	5.0	4.0	0.103	0.422
Wisconsin card sorting test perseverative responses	22.5	25.5	0.930	-0.023
Wisconsin card sorting test perseverative errors	20.0	21.75	0.956	-0.014
Wisconsin card sorting test non-perseverative errors	14.0	14.0	0.665	-0.121
Wisconsin card sorting test perseverative error percentage	17.09	16.75	0.903	0.033
Wisconsin card sorting test first category completion trials	13.0	9.0	0.786	0.076
Wisconsin card sorting test conceptual level responses	68.0	34.75	0.748	0.087
Wisconsin card sorting test conceptual level response percentage	53.45	31.68	0.557	0.158
Failure to maintain set score	1.0	2.0	0.059	-0.481
9-hole peg test placement (dominant hand)	16.0	6.25	0.071	-0.448
9-hole peg test placement (non-dominant hand)	15.5	7.58	0.106	-0.404
9-hole peg test removal (dominant hand)	9.0	4.56	0.066	-0.454
9-hole peg test removal (non-dominant hand)	8.5	4.75	**0.046**	-0.488
25-foot walk test	16.0	11.75	0.967	-0.010

Spearman correlation test was used for determination of p-values
MMSE: Mini-mental state examination, ACE-R : Addenbrooke's Cognitive Examination-Revised, WMS-R: Wechsler Memory Scale-Revised, COWAT: Controlled Oral Word Association Test, CVLT: California Verbal Learning Test, NBD: Neuro-Behçet's disease, PADI4: Peptidylarginine deiminase type 4

suggest that PADI4 serves as a general marker of immune activation and inflammation rather than a disease-specific indicator, which limits its utility for differential diagnosis.

Furthermore, the lack of association with clinical and cognitive measures diminishes the potential of PADI4 as a prognostic biomarker. This is particularly relevant considering that NBD presents with a relapsing-remitting course and variable neurological manifestations, often resolving spontaneously (3). Because all patients in our study were evaluated during remission, it remains unclear whether PADI4 expression fluctuates with disease activity and correlates with measures of disability during the attack period. Future studies should compare patients during both the attack and remission phases to better understand whether PADI4 could serve as a useful biomarker for disease activity and progression in NBD.

Intriguing questions raised by our study are why PADI4 levels are elevated in inflammatory disorders of the brain and how elevated PADI4 levels might affect disease mechanisms. It is plausible that PADI4 and NETosis are involved in divergent disease mechanisms across autoimmune and inflammatory disorders. For instance, in MS, PADI4 has been implicated in the citrullination of myelin basic protein, potentially contributing to increased immunogenicity of myelin proteins and triggering immune responses against the myelin tissue (12). In a few studies focused on the association between NETosis and BD, PADI4 and NETosis have been implicated to contribute to disease occurrence by increasing neutrophil activity and thereby neutrophil-mediated tissue damage and by inducing a propensity to thrombosis through endothelial damage, platelet activation, increased coagulation, and thrombin generation (13-15). As a matter of fact, neutrophils from BD patients are more prone to NETosis even in the absence of stimuli (13). Likewise, the intimate association between neutrophil activity and clinical severity in patients with NMO and NMO spectrum disorders might explain increased PADI4 levels in these disorders. Indeed, PADI4 levels are not only increased in NMO patients but also sharply decline following immunotherapy in parallel to clinical amelioration, further corroborating the link

between NETosis and inflammation in autoimmune brain disorders (16).

Peptidylarginine deiminase type 4's function in NET formation and neutrophil activity may help explain its elevation not only in NBD (5,6,8) but also in patients with a positive pathergy test, which indicates a link between PADI4 and exaggerated innate immune responses. This finding supports the idea that PADI4 may contribute to the increased tissue innate immune system reactivity manifesting itself as a positive pathergy test, a hallmark feature of BD (6).

Cognitive impairment involving attention, memory, and executive function has been frequently reported in both BD and NBD (3). Another intriguing question was whether NETosis-mediated mechanisms could be involved in the derangement of cognitive and motor functions in NBD. As an example, in MS, PADI4-mediated tissue alterations have been suggested to contribute to neuroinflammatory processes that can impair motor function in due course (12). This underscores the need for studies exploring whether NETosis-related pathways contribute to specific cognitive or motor outcomes across immune-mediated diseases. Contrary to this notion, our study failed to show a genuine association between PADI4 levels and cognitive and somatic dysfunction observed in NBD patients. Nevertheless, additional markers of NETosis or neutrophil activity, particularly when assessed in peripheral blood or cerebrospinal fluid, might show an association with neuropsychological features of NBD and thus warrant exploration in future studies.

Study Limitations

This study has several limitations, including a relatively small sample size and a cross-sectional design, which constrain the ability to infer causal relationships or monitor temporal changes in PADI4 levels. By focusing solely on patients in remission, the study may have overlooked potential fluctuations associated with active disease phases. The lack of disease specificity of PADI4, as it is also elevated in other autoimmune disorders, limits its diagnostic utility. Additionally, the trend-level correlations with cognitive measures and the limited neuropsychological assessment may have failed to detect more nuanced effects. Despite these limitations, our study has shown for the first time the potential significance of PADI4, a well-known marker of NETosis, in NBD and thus has provided a rationale for future larger longitudinal studies that might further clarify our findings.

Conclusion

This study highlights PADI4 as a potential contributor to immune dysregulation in NBD, reinforcing the putative role of NETosis in NBD and other inflammatory brain disorders. However, its lack of disease specificity and limited correlation with clinical features argue against the value of serum PADI4 as a prognostic and diagnostic marker in NBD. Nevertheless, assessment of PADI4 in peripheral blood and central nervous system tissues in future studies might be more useful for uncovering PADI4- and NETosis-related mechanisms in NBD.

Ethics

Ethics Committee Approval: Ethical approval was taken from the Istanbul University Faculty of Medicine Clinical Research Ethics Committee (approval no: 09, date: 16.04.2021).

Informed Consent: Informed consent was obtained from patients for examination of their blood samples for research purposes.

Footnotes

Authorship Contributions

Concept: C.I.K., M.K., Design: C.I.K., T.G., M.K., Data Collection or Processing: N.S., M.T., Analysis or Interpretation: C.I.K., N.S., H.Y.K., T.G., M.K., Literature Search: N.S., M.T., H.Y.K., Writing: C.I.K., N.S.

Conflict of Interest: The authors declared that there were no conflicts of interest.

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

References

1. Zhan H, Cheng L, Li Y. Neuro-Behçet's disease: an update of clinical diagnosis, biomarkers, and immunopathogenesis. Clin Exp Immunol. 2025;219:uxae123.
2. Al-Omoush O, AlBarakat MM, Alasmar D, et al. Clinical and imaging features of parenchymal and non-parenchymal neuro-Behçet's disease: a systematic review of case reports and series. BMC Neurol. 2025;25:225.
3. Camard M, Urbain F, Noel N. Cognitive impairment in systemic autoimmune and inflammatory diseases: a narrative review focused on ANCA-associated vasculitis, sarcoidosis, Sjögren's syndrome, systemic sclerosis, and Behçet's disease. Autoimmun Rev. 2025;24:103899.
4. Belfeki N, Ghriss N, Le Joncour A, Saadoun D. Etiopathogenesis of Behçet's disease: a systematic literature review. Clin Immunol. 2025;279:110549.
5. Bektas E, Deniz R, Emrence Z, et al. The relationship between NETosis findings and disease activity in Behçet's disease: an exploratory study. Clin Exp Rheumatol. 2025;43:1772-81.
6. Altan Ferhatoğlu Z, Altinişik DD, Özdede A, et al. Folliculitis might be associated with pathergy-positivity in patients with Behçet syndrome. Medicine (Baltimore). 2024;103:e37553.
7. Araujo-Abad S, García-Peñarrubia B, Giudici AM, et al. The citrullinating enzyme PADI4 binds to lipids: ıdentification of new target interactions for cancer therapy. J Mol Biol. 2025;437:169297.

8. Seol SI, Oh SA, Davaanyam D, Lee JK. Blocking peptidyl arginine deiminase 4 confers neuroprotective effect in the post-ischemic brain through both NETosis-dependent and -independent mechanisms. Acta Neuropathol Commun. 2025;13:33.
9. Quan M, Zhang H, Deng X, Liu H, Xu Y, Song X. Neutrophils, NETs and multiple sclerosis: a mini review. Front Immunol. 2025;16:1487814.
10. Won T, Naik P, Wood MK, et al. Anti-peptidylarginine deiminase 4 autoantibodies derived from patients with rheumatoid arthritis exert pathogenic effects by activating monocytes and exacerbating inflammatory arthritis. Arthritis Rheumatol. 2025;77:1150-65.
11. Zheng B, Cai P, Chen Y, Wu Y, Liu H. PADI4 and IL-33 gene polymorphisms with susceptibility to systemic lupus erythematosus and juvenile idiopathic arthritis, a systematic review and meta-analysis. Medicine (Baltimore). 2023;102:e33700.
12. Kijak-Boćkowska M, Czerwińska J, Owczarczyk-Saczonek A. Peptidylarginine deiminases: an overview of recent advances in citrullination research. Int J Mol Sci. 2025;26:12060.
13. Batu ED. Neutrophil-mediated thrombosis and NETosis in Behçet's disease: a hypothesis. J Korean Med Sci. 2020;35:e213.
14. Zhao F, Guo Y, Jiao Y, et al. Mitochondrial metabolism in neutrophils: emerging roles in ınflammatory skin diseases. Br J Dermatol. 2026:ljag060.
15. Li P, Ji Y, Shen D, et al. Integrated analysis of intersecting neutrophil signatures in Behçet's disease and inflammatory bowel disease. Int J Rheum Dis. 2025;28:e70229.
16. Zhang W, Cui L, Jiao J, et al. Decreased NETosis-related regulators in neuromyelitis optica spectrum disorders after plasma exchange. Int Immunopharmacol. 2024;142:113234.