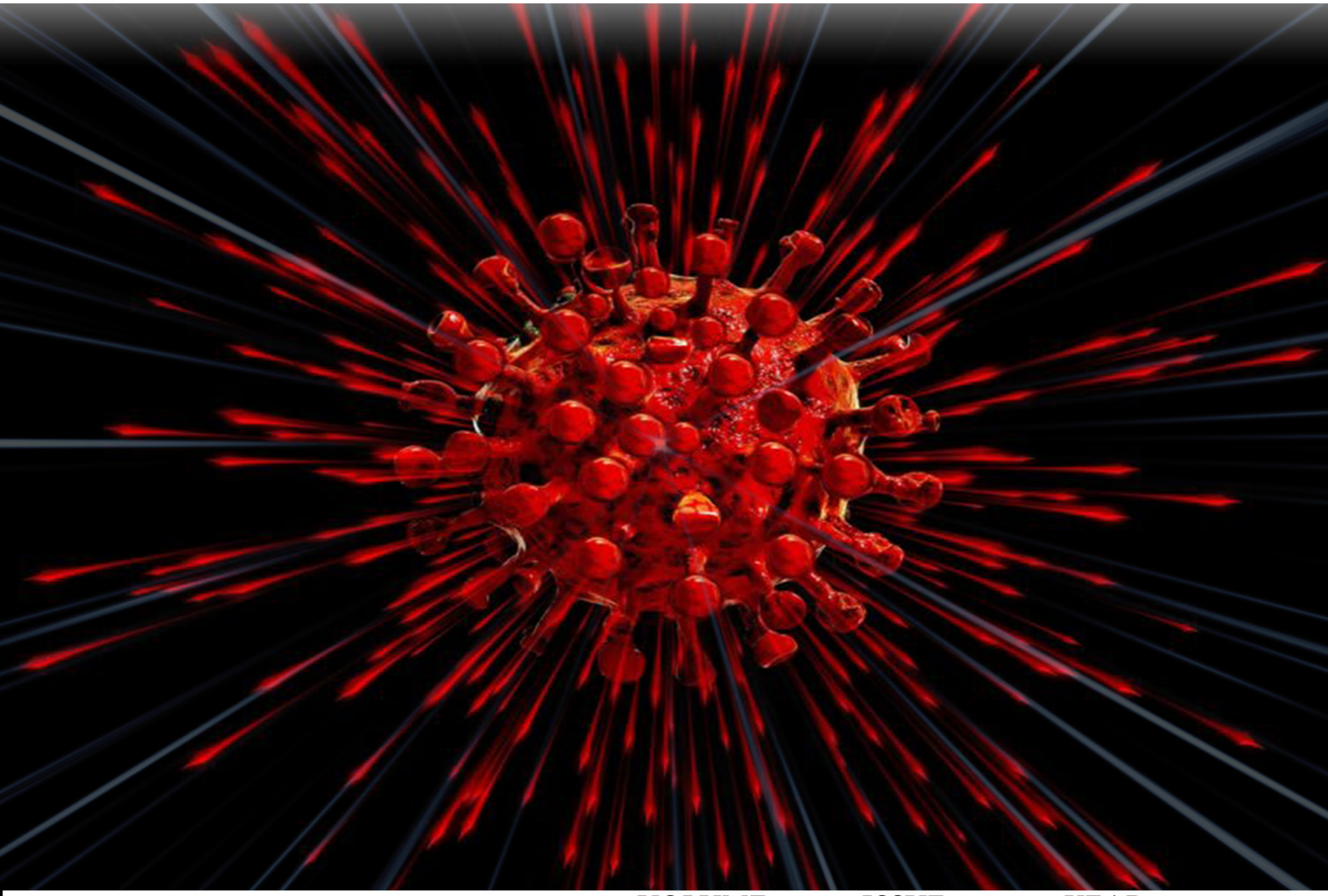
e-ISSN: 2718-0115

ACMJ

Anatolian Current Medical Journal



VOLUME: 7 ISSUE: 5 YEAR: 2025

MHA
MEDIHEALTH ACADEMY

CORRESPONDENCE ADDRESS

Sağlık Bilimleri Üniversitesi Teknoloji Geliştirme Merkezi (SBÜ TEKMER)
Gülhane Yerleşkesi, Emrah Mah. General Dr. Tevfik Sağlam Cad. No: 2/67
İç Kapı No: 105, Keçiören/Ankara/Turkiye
E-mail: mha@medihealthacademy.com

www.medihealthacademy.com

ARTICLE SUBMISSION ADDRESS

https://dergipark.org.tr/tr/journal/2316/submission/step/manuscript/new

EDITORIAL BOARD

CONTENTS

Original Articles

CONTENTS

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1707632

Anatolian Curr Med J. 2025;7(5):546-550

Comparison of SVEAT and HEART scoring in acute chest pain

Aynur Yurtseven[1], Cemil Kavalcı[2], Afşin Emre Kayıpmaz[1]

[1]*Department of Emergency Medicine, Ankara Etlik City Hospital, Ankara, Turkiye*

[2]*Department of Emergency Medicine, Antalya Training and Research Hospital, Antalya, Turkiye*

Cite this article as: Yurtseven A, Kavalcı C, Kayıpmaz AE. Comparison of SVEAT and HEART scoring in acute chest pain. *Anatolian Curr Med J.* 2025;7(5):546-550.

Received: 27.05.2025 • **Accepted**: 22.07.2025 • **Published**: 15.09.2025

ABSTRACT

Aims: It is very important to evaluate patients presenting with chest pain in terms of major adverse cardiovascular events (MACE) and many risk scoring systems have been developed for this purpose. In this study, we aimed to evaluate the MACE prediction performance of the newly developed symptoms, history of vascular disease, electrocardiography, age and troponin (SVEAT) score for patients presenting with chest pain.

Methods: This study was designed as a retrospective observational clinical trial. MACE occurring within 30 days; Myocardial infarction (MI), percutaneous coronary intervention (PCI), coronary artery bypass graft (CABG) and sudden cardiac death were considered the primary endpoints of the study. Patients over 18 years of age presenting with chest pain were included in the study. Patients with ST segment elevation on electrocardiography (ECG), hemodynamic instability and traumatic chest pain were excluded.

Results: The study included 557 patients and the mean age was 54.52±12.56 and the age distribution range was 23-95. Significant results for SVEAT score (AUC:0.988, 95%CI:0.978-0.997, p<0.001) and history, electrocardiography, age, risk factors and troponin (HEART) score (AUC:0.960, 95%CI:0.942-0.979, p<0.001) were obtained from ROC analysis of the effect of SVEAT and HEART scores on MACE.

Conclusion: The newly developed SVEAT score was superior to the HEART score in predicting adverse negative cardiac events in patients presenting with chest pain.

Keywords: Chest pain, emergency, risk score, cardiovascular event

INTRODUCTION

Chest pain is one of the most common reasons for being admitted to the emergency department. Clinically, acute chest pain may be associated with a clinical condition with a high mortality rate such as myocardial infarction (MI), pulmonary thromboembolism, aortic dissection; it may also be associated with musculoskeletal diseases, pain reflected from internal organs such as gastrointestinal tract or sometimes psychiatric causes.[1]

Coronary artery disease (CAD) is now one of the most common causes of mortality and morbidity. In the United States, it is estimated that approximately 1.0 million people present with acute chest pain annually and 300.000 to 400.000 people die due to MI.[2] Studies show that cardiovascular diseases account for 45% of deaths in women and 38% of deaths in men under the age of 75 in Europe.[3,4] Heart and artery diseases are predicted to be staying as the top cause of death for a long time as a result of the expected increase in the estimated life expectancy.[5]

Acute coronary syndrome (ACS) must be quickly differentiated from other clinically similar conditions. Studies have shown that approximately 2-4% of patients with undetected MI on initial presentation are wrongly discharged, accounting for 20-39% of all emergency malpractice cases.[6,7] Therefore, early diagnosis of ACS in patients defined as low risk group is of critical importance. Many risk scoring systems have been developed in this patient group in order to start treatment with early diagnosis and to decrease the median mortality related to ACS.

The usability of the risk scoring system can be evaluated by its performance in detecting the possibility of developing MACE in the early period.[8] The Thrombolysis in Myocardial Infarction (TIMI) score,[9] the Global Registry of Acute Coronary Events (GRACE)[10] and the History, Electrocardiography, Age, Risk Factors and Troponin (HEART) generation HEART score[11] are risk scoring systems used in patients presenting with chest pain. However, there is no current consensus on which of these scoring systems is more reliable for the emergency department.[12]

Symptoms, Vascular Disease History, Electrocardiography, Age and Troponin (SVEAT) risk score, presented as a slightly

Corresponding Author: Aynur Yurtseven, aynuryurt7@gmail.com

better performance than the currently widely used HEART score.[13] SVEAT scoring system, in the case of an acute attack event, higher acceleration, negative scoring in the event of no event and a much wider scoring range. The ratios of the clinical risk composite were seen to increase with a wider score range.[14] In addition, a change in the ECG in the SVEAT scoring was defined with much sharper boundaries and the characteristics of chest pain were considered in detail and vascular diseases were also included in the score.[14] This study aimed to compare SVEAT and HEART scores in terms of MACE predictability in patients presenting to the emergency department with chest pain.

METHODS

Ethics

The study followed the principles of the Declaration of Helsinki revised in 2013. This study has been approved by the Ankara Etlik City Hospital No. 1 Local Ethics Committee for Clinical Researches (Date: 01.11.2023, Decision No: AEŞH-EK1-655-2023).

Study Design and Participants

The study consisted of patients who were admitted to the city hospital emergency department with complaints of acute chest pain between 01.01.2022 and 31.12.2023. Patients who had not previously been used for either scoring system (SVEAT and HEART) for patient outcomes were included in the study. Patients with hemodynamic instability, acute MI detected on admission ECG, those presenting with chest pain due to trauma, and those with incomplete medical records were excluded from the study. A total of 557 patient files from 1162 patients presenting with chest pain who met the study criteria were analyzed (**Figure 1**). The number of patients to be included in the study was determined using the G-power 3.1.9.4 program with a power of 80% and a significance level of p<0.05.

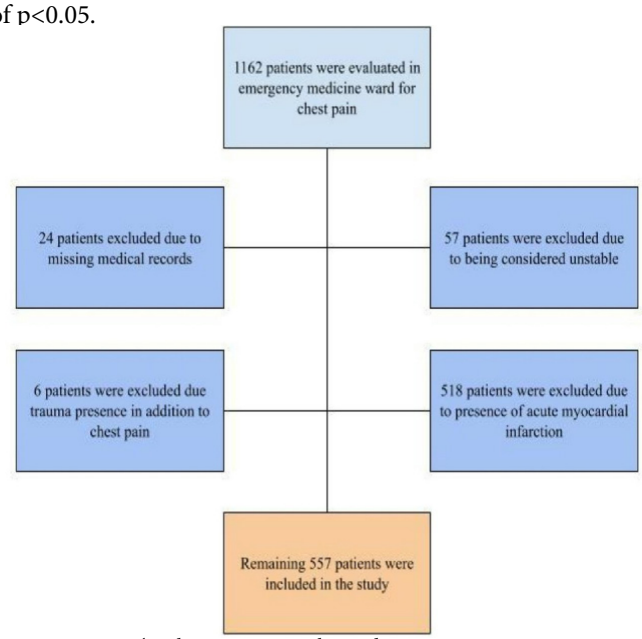


Figure 1. Patients' inclusion status in the study

Troponin level was obtained from the data of the fourth generation ultra-high sensitivity troponin assay (Elecysd Roche Kit) and determined as the upper normal limit Troponin (TnT-hs) level equivalent to 14 ng/L (pg/ml).

Data Collection

The risk scores of all patients were calculated according to SVEAT and HEART scores. The parameters required for SVEAT scoring are given in **Table 1**. In addition, age, gender and underlying medical comorbidities (diabetes, hypertension, intracranial hemorrhage and stroke) were recorded. MI requiring revascularization or medical treatment and sudden cardiac arrest MACE occurring after readmission within 30 days were accepted as the study endpoints.

Table 1. SVEAT scoring system

Component	Characteristics	Points
Symptoms	Typical unstable angina pectoris	3
	Stable angina, Canadian Cardiovascular Society Class I or II	1
	Non-cardiac chest pain	-2
Vascular disease	Recent myocardial infarction or percutaneous coronary intervention <90 days	2
	Coronary artery bypass grafting >5 years	2
	Prior coronary event other than above	1
	Prior revascularization for peripheral disease or carotid disease	2
ECG	Dynamic or new ischemic ST or T wave changes	3
	ST depression of unknown duration without cause	2
	ST changes with left ventricular hypertrophy, intraventricular conduction delay, digitalis, or metabolic issue	1
	Old Q wave indicating prior myocardial infarction or pre-existing ST changes	1
	No ST changes	0
	Normal ECG in the presence of severe ongoing chest pain	-2
Age (years)	>75	2
	50-75	1
	30-49	0
	<30	-1
Troponin I (ng/ml)	0.7 or higher	5
	>0.12 but <0.7	2
	2>0.04 but <or=0.12	1
	Normal (<or=0.004) with unclear duration of chest pain	0
	Normal after >4 h of constant chest pain	-2

SVEAT: Symptoms, Vascular Disease History, Electrocardiography, Age and Troponin, ECG: Electrocardiography

Statistical Analysis

SPSS 20.0 for Windows® statistical software (IBM Inc. Chicago, IL, USA) was used for statistical analysis. Number

(n), percentage (%), mean, standard deviation, median, minimum and maximum values were used in the presentation of descriptive data. The conformity of the data to normal distribution was assessed using the Kolmogorov-Smirnov test. Pearson Chi-square test and Fisher's exact test were used for comparison of categorical data, T test for comparison of two independent numerical data, Mann-Whitney U test for unequal distribution of data, Kruskal-Walles Test and ANOVA tests were used for comparison of two numerical data. ROC analysis of the area under the curve was used to compare the predictability of MACE of SVEAT and HEART scores. p<0.05 and 95% confidence interval were accepted for statistical significance.

RESULTS

The files of 1162 patients who presented with chest pain were reviewed and 557 patients were included in the study. The mean age of the patients was 54.52±12.56 years and the age distribution was 23-95. 233 (41.7%) of the patients included in the study were female. Mortality was observed in only 3 (0.5%) of the patients included in the study.

MACE was positive in 110 patients and MACE was negative in 447 patients. When the factors affecting MACE were examined, advanced age, hypertension, diabetes, history of MI, percutaneous coronary intervention (PCI), coronary artery bypass grafting (CABG), low SVEAT and HEART scores were statistically significant (p<0.001). Similarly, the effect of hyperlipidemia history on MACE was also found to be significant (p:0.016). Gender and stroke history had no significant effect on MACE (**Table 2**). Significant results were obtained in ROC analysis for SVEAT (AUC:0.988, 95% Cl:0.978-0.997, p<0.001) and HEART (AUC:0.960, 95% Cl:0.942-0.979, p<0.001) (**Table 3, Figure 2**).

Table 2. MACE distribution based on risk factors

	MACE (-) (n:447)	MACE (+) (n:110)	p value
Age	52.66±12.04	62.09±11.88	<0.001
Female sex	197	35	0.055
Hyperlipidemia	121	42	0.016
Hypertension	149	71	<0.001
Diabetes	99	51	<0.001
Stroke	9	2	0.626
MI history	108	68	<0.001
PCI	96	63	<0.001
CABG	16	14	<0.001
SVEAT	(-) 0.32±2.04	8.41±3.07	<0.001
HEART	2.40±1.67	6.85±1.68	<0.001

MACE: Major adverse cardiovascular events, MI: Myocardial infarction, PCI: Percutaneous coronary intervention, CABG: Coronary artery bypass grafting, SVEAT: Symptoms, Vascular Disease History, Electrocardiography, Age and Troponin, HEART: History, Electrocardiography, Age, Risk Factors and Troponin

The mean scores of SVEAT and HEART were calculated as 1.40±4.16 and 3.28±2.44, respectively. The best thresholds for SVEAT and HEART were determined as 3.5 and 4.5, respectively. While the sensitivity and specificity for SVEAT were 96.4% and 97.3%, respectively; these rates were found as 93.6% sensitivity and 87.5% selectivity for HEART. According to the best threshold values, the accuracy values for SVEAT and HEART were calculated as 541/557 (97.1%) and 494/557 (88.7%), respectively (**Table 4**). These results showed that the SVEAT score was significant in detecting major adverse cardiovascular events (MACE). It was observed that MACE was missed by not being predicted in only 16 patients at the value determined as the best SVEAT threshold value. On the other hand, MACE unpredictability was found in 63 patients in HEART scoring. According to the study results, it was determined that the SVEAT scoring had better sensitivity and specificity values than the HEART score.

Table 3. SVEAT ve HEART scoring based on ROC analysis

Test result variable(s)	Under the curve	Std. error	p value	95% confidence interval	
				Lower bound	Upper bound
HEART	0.960	0.010	<0.001	0.942	0.979
SVEAT	0.988	0.005	<0.001	0.978	0.997

SVEAT: Symptoms, Vascular Disease History, Electrocardiography, Age and Troponin, HEART: History, Electrocardiography, Age, Risk Factors and Troponin, ROC: Receiver operating characteristic

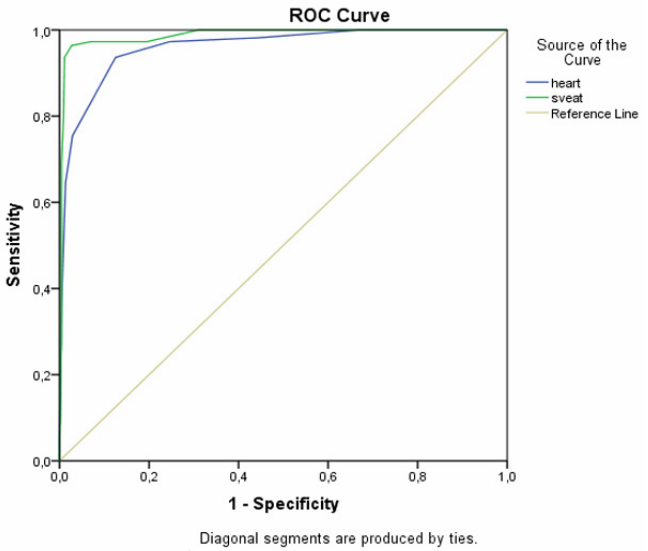


Figure 2. ROC Curve of SVEAT and HEART scoring
SVEAT: Symptoms, Vascular Disease History, Electrocardiography, Age and Troponin, HEART: History, Electrocardiography, Age, Risk Factors and Troponin, ROC: Receiver operating characteristic

Table 4. MACE prediction of the SVEAT and HEART scoring

Parameters	Definitions	SVEAT	HEART
Under the curve		0.988	0.96
95% confidence interval		0.978-0.997	0.942-0.979
p-value		<0.001	<0.001
The best threshold		>3.5	>4.5
Event size	N	557	557
Sensitivity	TP/(TP+TN)	106/110 (96.4%)	103/110 (93.6%)
Selectivity	TN/(TN+FP)	435/447 (97.3%)	391/447 (87.5%)
PPV	TP/(TP+FP)	106/118 (89.8%)	103/159 (64.8%)
NPV	TN/(FN+TN)	435/439 (99,1%)	391/398 (98.2%)
Truth	(TP+TN)/(N)	541/557 (97.1%)	494/557 (88.7%)
p value		<0.001	<0.001

MACE: Major adverse cardiovascular events, SVEAT: Symptoms, Vascular Disease History, Electrocardiography, Age and Troponin, HEART: History, Electrocardiography, Age, Risk Factors and Troponin, TP: True positive, FN: False negative, TN: True negative, FP: False positive, PPV: Positive predicted value, NPV: Negative predicted value

DISCUSSION

Early identification of risk groups in patients presenting to the hospital with chest pain may be the most important point for predicting MACE, and many different risk scoring systems have been developed for this purpose. Although the HEART score was developed for use in emergency departments for patients presenting with chest pain and is the most widely used risk score in the United States, there are studies in the literature showing that this scoring is not sufficient for the low-risk patient group.[15-17]

In the HEART score, patients are divided into low, medium and high risk groups according to the risk of developing MACE. It is recommended that those in the low risk group be discharged and those with medium and high risk scores be hospitalized.[18] In our study, the sensitivity of the HEART score of 93.6% and the specificity of 87.5% were found to be consistent with previous studies in the literature.[16,18,19] However, it was observed that the classification of the risk into high, medium and low risk groups in the distinction of chest pain varied depending on the evaluator. In addition, it was noted that some important cardiac clinical information in the initial evaluation was not fully used in this scoring and especially that significant ST depression was not clearly defined in the scoring.

A new scoring system (SVEAT score), developed by Roongsritong et al.[13] and based on five clinical variables: chest pain symptom characteristics, history of vascular disease, electrocardiography, age, and troponin, has been reported to outperform the HEART score. Unlike the HEART and TIMI risk scores, the SVEAT score assigns a 5-point value to symptoms suggestive of ischemia and a negative value to non-cardiac chest pain to better distinguish between subgroups. Furthermore, this scoring system uses the presence of cardiovascular disease rather than traditional risk factors and focuses more specifically on possible electrocardiographic (ECG) changes. It also assigns a higher score for troponin levels, resulting in a wider score range. This is thought to aid clinicians in their decision-making process. A limited number of studies in the literature have shown that the SVEAT score outperforms the HEART score.[13]

Roongsritong et al.[13] compared the SVEAT score with HEART and TIMI scores in a prospective study of 321 patients. The calculated AUC values were found to be 0.982, 0.921 and 0.884 for the SVEAT score, HEART score and TIMI score, respectively. Male gender, hypertension, diabetes mellitus and hyperlipidemia are seen as statistically significant determinants for the 30-day MACE score. Similarly, in our study, the AUC value of the SVEAT score was found to be higher than the HEART score, and the obtained results support Roongsritong's[13] study.

In the Antwi-Amoabeng[20] study, the threshold value was taken as ≤4 and the SVEAT score was compared with the HEART score, and it was concluded that the SVEAT score is a reliable predictor of cardiovascular morbidity. Similarly, we determined the threshold values as 3.5 for SVEAT and 4.5 for HEART. With these results, it was concluded that the SVEAT score has statistical significance in predicting MACE.

In the prospective study of Shahid et al.,[21] which included 60 patients and compared the SVEAT, HEART and TIMI risk scores, the sensitivity and specificity for the SVEAT and HEART scores were found to be 63.2%, 84.2% and 75.6%, 73.2%, respectively, at similar threshold values for SVEAT and HEART scores, and it was concluded that the SVEAT score was not superior to the HEART score in predicting 30-day MACE. In our study, 96.4% sensitivity and 97.3% selectivity were found for SVEAT, and 93.6% sensitivity and 87.5% selectivity were found for HEART, which is consistent with the original SVEAT study and the limited number of studies in the literature. This difference found by Shahid et al.[21] may be due to the study including only a very limited patient population aged 45 and over. In addition, unlike this study, we found that male gender was more dominant and stroke did not have a significant effect on MACE.

In this study, factors affecting the MACE parameter in hospitalized patients were found to be statistically significant, including advanced age, hypertension, diabetes, history of hyperlipidemia, current MI, PCI, history of CABG, high SVEAT and HEART scores, and were consistent with the findings in the original SVEAT study.[13,20]

Limitations

The study was conducted in a single tertiary care center with a specific sample size and the SVEAT score was compared only with the HEART score, which may be limitations of the study. It is thought that studies using other risk scoring systems and conducted on a larger population will contribute to the literature.

CONCLUSION

Based on the results of this study, we suggest that the SVEAT scoring system outperforms the HEART score in predicting the risk of MACE in patients presenting to the emergency department with chest pain. Further studies with diverse populations and larger sample sizes are needed before the SVEAT score can be widely applied in risk stratification.

ETHICAL DECLARATIONS

Ethics Committee Approval

This study has been approved by the Ankara Etlik City Hospital No. 1 Local Ethics Committee for Clinical Researches (Date: 01.11.2023, Decision No: AEŞH-EK1-655-2023).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Chang AM, Fischman DL, Hollander JE. Evaluation of chest pain and acute coronary syndromes. *Cardiol Clin.* 2018;36(1):1-12. doi:10.1016/j.ccl.2017.08.001
2. Benjamin EJ, Virani SS, Callaway CW, Chamberlain AM, Chang AR, Cheng S. Heart disease and stroke statistics—2018 update: a report from the American Heart Association. *Circulation.* 2018;137(12):e67-e492. doi:10.1161/CIR.0000000000000558
3. Perk J, De Backer G, Gohlke H, Graham I, Reiner Ž, Verschuren M. European guidelines on cardiovascular disease prevention in clinical practice (version 2012). *Eur Heart J.* 2012;33(13):1635-1701. doi:10.1093/eurheartj/ehs092
4. SCORE2 working group and ESC Cardiovascular risk collaboration. SCORE2 risk prediction algorithms: new models to estimate 10-year risk of cardiovascular disease in Europe. *Eur Heart J.* 2021;42(25):2439-2454. doi:10.1093/eurheartj/ehab309
5. Türkiye Kalp ve Damar Hastalıkları Önleme ve Kontrol Programı 2015–2020 Klavuzu. Ankara: Anıl Matbaa Ltd. Şti; 2005. T.C. Sağlık Bakanlığı Yayın No: 988.
6. Fuller M, Hamilton D, Holly J, Mallin M, Rayner T, Eshenroder N. Prospective evaluation of a simplified risk stratification tool for patients with chest pain in an emergency department observation unit. *Crit Pathw Cardiol.* 2013;12(3):132-136. doi:10.1097/HPC.0b013e31829a79cd
7. Halpern EJ, Deutsch JP, Hannaway MM, Estepa AT, Kenia AS, Neuburger KJ. Cardiac risk factors and risk scores vs cardiac computed tomography angiography: a prospective cohort study for triage of ED patients with acute chest pain. *Am J Emerg Med.* 2013;31(10):1479-1485. doi:10.1016/j.ajem.2013.08.001
8. Akyea RK, Leonardi-Bee J, Asselbergs FW, Patel RS, Durrington P, Wierzbicki AS. Predicting major adverse cardiovascular events for secondary prevention: protocol for a systematic review and meta-analysis of risk prediction models. *BMJ Open.* 2020;10(7):e034564. doi:10.1136/bmjopen-2019-034564
9. Antman EM, Cohen M, Bernink PJ, et al. The TIMI risk score for unstable angina/non–ST elevation MI: a method for prognostication and therapeutic decision making. *JAMA.* 2000;284(7):835-842. doi:10.1001/jama.284.7.835
10. Eagle KA, Lim MJ, Dabbous OH, et al. A validated prediction model for all forms of acute coronary syndrome: estimating the risk of 6-month postdischarge death in an international registry. *JAMA.* 2004;291(22):2727-2733. doi:10.1001/jama.291.22.2727
11. Poldervaart JM, Reitsma JB, Koffijberg H, et al. The impact of the HEART risk score in the early assessment of patients with acute chest pain: design of a stepped wedge, cluster randomised trial. *BMC Cardiovasc Disord.* 2013;13(1):1-8. doi:10.1186/1471-2261-13-77
12. Sakamoto JT, Liu N, Koh ZX, et al. Comparing HEART, TIMI, and GRACE scores for prediction of 30-day major adverse cardiac events in high acuity chest pain patients in the emergency department. *Int J Cardiol.* 2016;221:759-764. doi:10.1016/j.ijcard.2016.07.147
13. Roongsritong C, Taha ME, Pisipati S, et al. SVEAT score, a potential new and improved tool for acute chest pain risk stratification. *Am J Cardiol.* 2020;127:36-40. doi:10.1016/j.amjcard.2020.04.009
14. Hollander JE, Than M, Mueller C. State-of-the-art evaluation of emergency department patients presenting with potential acute coronary syndromes. *Circulation.* 2016;134(7):547-564. doi:10.1161/CIRCULATIONAHA.116.021886
15. Backus BE, Six AJ, Kelder JC, et al. A prospective validation of the HEART score for chest pain patients at the emergency department. *Int J Cardiol.* 2013;168(3):2153-2158.1 doi:10.1016/j.ijcard.2013.01.255
16. Backus BE, Six AJ, Kelder JC, et al. Chest pain in the emergency room: a multicenter validation of the HEART score. *Crit Pathw Cardiol.* 2010;9(3):164-169. doi:10.1097/HPC.0b013e3181ec36d8
17. Poldervaart JM, Langedijk M, Backus BE, et al. Comparison of the GRACE, HEART and TIMI score to predict major adverse cardiac events in chest pain patients at the emergency department. *Int J Cardiol.* 2017;227:656-661. doi:10.1016/j.ijcard.2016.10.080
18. Van den Berg P, Body R. The HEART score for early rule out of acute coronary syndromes in the emergency department: a systematic review and meta-analysis. *Eur Heart J Acute Cardiovasc Care.* 2018;7(2):111-119. doi:10.1177/2048872617710788
19. Mahler SA, Hiestand BC, Goff DC Jr, Hoekstra JW, Miller CD. Can the HEART score safely reduce stress testing and cardiac imaging in patients at low risk for major adverse cardiac events? *Crit Pathw Cardiol.* 2011;10(3):128-133. doi:10.1097/HPC.0b013e3182315a85
20. Antwi-Amoabeng D, Roongsritong C, Taha M, et al. SVEAT score outperforms HEART score in patients admitted to a chest pain observation unit. *World J Cardiol.* 2022;14(6):454-461. doi:10.4330/wjc.v14.i8.454
21. Shahid MF, Malik A, Kashif N, Siddiqi FA, Hammad M, Saeed HA. Risk stratification of acute-onset chest pain: SVEAT score versus HEART and TIMI scores. *Cureus.* 2023;15(5):e39590. doi:10.7759/cureus.39590

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1720706

Anatolian Curr Med J. 2025;7(5):551-558

The amplifying effect of maternal obesity on perinatal outcomes in gestational diabetes mellitus

Erkan Sağlam

Department of Perinatology, Bursa City Hospital, Bursa, Turkiye

Cite this article as: Sağlam E. The amplifying effect of maternal obesity on perinatal outcomes in gestational diabetes mellitus. *Anatolian Curr Med J.* 2025;7(5):551-558.

Received: 16.06.2025 ◆ **Accepted**: 22.07.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: This study aimed to evaluate the impact of maternal body-mass index (BMI) in the first trimester on perinatal outcomes in pregnant women diagnosed with gestational diabetes mellitus (GDM).

Methods: This retrospective cohort study included 210 women with GDM, categorized into three groups based on first-trimester BMI: normal weight (BMI<25.0 kg/m²), overweight (BMI 25.0–29.9 kg/m²), and obese (BMI≥30.0 kg/m²). Perinatal outcomes such as preterm birth, low birth weight, macrosomia, NICU admission, and Apgar scores were compared across groups. Multivariate logistic regression analyses were conducted to evaluate the independent association between obesity and adverse perinatal outcomes.

Results: The rates of preterm birth <37 weeks (35.7%, p<0.001) and <34 weeks (11.4%, p=0.016) were significantly higher in the obese group. Although low birth weight (<2500 g) was observed in 10.0% of obese women, this difference was not statistically significant (p=0.211). Macrosomia (≥4000 g) was more frequent in the obese group (11.4%) but did not reach statistical significance (p=0.261). NICU admission rates were higher in obese women (22.9%, p=0.089). In adjusted analyses, obesity was independently associated with increased odds of preterm birth <34 weeks (adjusted OR: 6.01, p=0.012) and low birth weight (adjusted OR: 4.68, p=0.026). Additionally, Spearman correlation analysis revealed a weak, non-significant positive correlation between gestational weight gain and gestational age at birth (r=0.122, p=0.077).

Conclusion: In pregnant women with GDM, first-trimester maternal obesity is associated with an increased risk of early preterm birth and low birth weight. These findings underscore the importance of assessing and managing maternal BMI early in pregnancy to improve perinatal outcomes.

Keywords: Gestational diabetes mellitus, maternal obesity, perinatal outcomes, preterm birth

INTRODUCTION

Variable levels of hyperglycemia that are initially identified during gestation are the hallmark of gestational diabetes mellitus (GDM), a common metabolic disease of pregnancy.[1] Between 13.97% and 14.8% of people worldwide have GDM, and its incidence is rising in tandem with rising obesity rates.[2,3]

One of the most important risk factors for the development of GDM is obesity, which is defined as having a pre-pregnancy body-mass index (BMI) of ≥30 kg/m². Compared to women of normal weight, those who are obese have a three to nine times higher risk of developing GDM.[4] The primary causes of this association include insulin resistance and persistent low-grade inflammation seen in obese people.[5,6]

Adverse prenatal outcomes, such as preterm birth, hypertensive problems, cesarean sections, and neonatal issues including low birth weight or macrosomia, are caused by both GDM and obesity on their own.[7-11] Furthermore, these risks may be increased if maternal obesity and GDM coexist.[12]

Although the synergistic consequences of obesity and GDM are becoming more well acknowledged, there are contradictions in the literature about which condition is more responsible for particular negative outcomes. These differences could be due to the confounding effect of excessive gestational weight gain (GWG), the timing of BMI assessment (first vs. second trimester), and differences in GDM screening techniques (universal vs. risk-based).[13-15]

The purpose of this study was to assess the relationship between pre-pregnancy BMI and unfavorable perinatal outcomes in women with GDM. We aimed to evaluate the independent impact of maternal obesity on obstetric and neonatal problems by utilizing first-trimester BMI data and controlling for GWG in multivariate models.

Corresponding Author: Erkan Sağlam, dresaglam@gmail.com

METHODS

This study has been approved by the Scientific Researches Ethics Committee of Bursa City Hospital (Date: 14.05.2025, Decision No: 2025/10-12). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

This retrospective cohort study was conducted at a tertiary care center between January 2022 and December 2024. Medical records of 1.128 pregnant women who delivered at our institution and had recorded BMI values between the 8th and 12th weeks of gestation were screened. After applying exclusion criteria (pre-gestational diabetes, chronic hypertension, multifetal gestation, and missing data), 345 patients with GDM diagnosed according to the International Association of Diabetes and Pregnancy Study Groups (IADPSG) criteria were identified. To obtain equal group sizes for comparison, 210 patients were selected and categorized into three groups of 70 each.

GDM diagnosis was made if at least one of the following thresholds was met during a 75-g oral glucose tolerance test (OGTT): fasting plasma glucose ≥92 mg/dl, 1-hour ≥180 mg/dl, or 2-hour ≥153 mg/dl. Based on their first trimester BMI, participants were categorized into three groups: normal weight (BMI<25.0 kg/m^2), overweight (BMI 25.0–29.9 kg/m^2), and obese (BMI≥30.0 kg/m^2).

Maternal BMI was calculated using anthropometric data recorded during the first antenatal visit between the 8th and 12th weeks of gestation. Maternal demographic and clinical characteristics including maternal age, gravidity, parity, smoking status, in vitro fertilization (IVF) history, and treatment modality (diet or insulin) were obtained from the hospital's electronic medical record system.

All patients received individualized dietary counseling from a certified dietitian according to national gestational diabetes guidelines. Insulin therapy was initiated in cases where glycemic targets were not achieved within two weeks of diet modification. Dose titration was tailored according to fasting and postprandial blood glucose levels.

Laboratory and clinical parameters recorded in the third trimester or at delivery were collected. These included GWG, fasting glucose level, HbA1c, total cholesterol, triglycerides, white blood cell count (WBC), hemoglobin (Hb), and platelet count.

Perinatal outcomes were extracted from delivery and neonatal records and included gestational age at delivery, birth weight, macrosomia (≥4000 g), preterm birth (<37 and <34 weeks), 1- and 5-minute Apgar scores, neonatal intensive care unit (NICU) admission, and mode of delivery (vaginal vs cesarean section).

Statistical Analysis

All data analyses were performed using IBM SPSS Statistics version 26.0. Descriptive statistics were used to summarize the demographic and clinical characteristics of the study population. Continuous variables were tested for normality using the Shapiro-Wilk test. Normally distributed variables were expressed as mean±standard deviation and compared using one-way ANOVA; non-normally distributed variables were expressed as median (minimum–maximum) and compared using the Kruskal-Wallis test. Categorical variables were expressed as numbers and percentages, and compared using the Chi-square or Fisher's exact test as appropriate.

Multivariate logistic regression models were constructed to assess the association between BMI categories and adverse perinatal outcomes, including preterm birth (<37 weeks and <34 weeks), low birth weight (<2500 g), macrosomia (≥4000 g), and NICU admission. Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated. Two models were used: model 1 was unadjusted; model 2 was adjusted for maternal age, smoking status, fasting glucose, HbA1c, treatment modality, and GWG. A p-value of <0.05 was considered statistically significant.

RESULTS

This study included a total of 210 pregnant women diagnosed with GDM between 2022 and 2024. All cases had complete data for GWG and perinatal outcomes. Patients were stratified into three groups based on their BMI measured at the first prenatal visit (8–12 gestational weeks):

- Normal weight: BMI 18.5–24.9 kg/m^2 (n=70)
- Overweight: BMI 25.0–29.9 kg/m^2 (n=70)
- Obese: BMI ≥30.0 kg/m^2 (n=70)

Maternal and Laboratory Characteristics

As shown in **Table 1**, although maternal age was higher in the obese group, the difference was not statistically significant (p=0.109). GWG was significantly lower in the obese group compared to the other two groups (p<0.001). Laboratory parameters including HbA1c, fasting glucose, WBC, and triglyceride levels were significantly elevated in the obese group (all p<0.001). Total cholesterol was also significantly higher in the obese group compared to the normal weight group, but not compared to the overweight group. Insulin therapy was more frequently required in obese (68.6%) and overweight (50.0%) women compared to normal weight (28.6%) women (p<0.001); however, the difference between obese and overweight groups was not statistically significant. Smoking status (p=0.402) and IVF-conceived pregnancy rates (p=0.218) did not differ significantly among the groups.

Perinatal Outcomes

As presented in **Table 2**, the rate of preterm delivery before 37 weeks was significantly higher in the obese group (35.7%) than in overweight (14.3%) and normal weight (11.4%) groups (p<0.001). Similarly, <34-week deliveries were more common in the obese group (11.4%, p=0.016). The median gestational age was significantly lower in the obese group (37.0 weeks, p<0.001). Birth weight did not differ significantly between groups (p=0.061), and the rate of low birth weight (<2500 g) was also not statistically significant (10.0% in obese group, p=0.211). The prevalence of macrosomia (birthweight ≥4000 g) was higher in the obese group (17.1%) than in overweight (10.0%) and normal weight (7.1%) groups, but the difference

Table 1. Demographic, clinical, and laboratory parameters by BMI group

Variable	BMI normal (n=70)	BMI overweight (n=70)	BMI obese (n=70)	p-value
Age (years)	29.00 (19.00–41.00)	30.00 (20.00–41.00)	31.00 (20.00–49.00)	0.109[1]
Gravida	2.00 (1.00–5.00)	2.00 (1.00–7.00)	2.00 (1.00–7.00)	0.180[1]
Abortus	0.00 (0.00–2.00)	0.00 (0.00–3.00)	0.00 (0.00–3.00)	0.445[1]
Smokers (%)	11 (15.7%)	10 (14.3%)	12 (17.1%)	0.898[2]
IVF-conceived pregnancies (%)	4 (5.7%)	3 (4.3%)	2 (2.9%)	0.706[2]
Treatment modality, n (%)				**<0.001[2]**
Diet	50 (71.4%)[a]	35 (50.0%)	22 (31.4%)[b]	
Insulin	20 (28.6%)[b]	35 (50.0%)	48 (68.6%)[a]	
GWG (kg)	13.0 (7–17)[a]	12.4 (7–16)[a]	9.4 (6–16)[b]	<0.001[1]
HbA1c (%)	5.04 (4.00–6.10)[c]	5.23 (4.69–6.50)[b]	5.65 (4.99–6.90)[a]	<0.001[1]
Fasting glucose (mg/dl)	92.50 (85.00–118.00)[c]	95.00 (83.00–119.00)[b]	98.00 (83.00–295.00)[a]	<0.001[1]
Hemoglobin (g/dl)	11.60 (8.30–14.00)	11.20 (7.60–14.00)	11.90 (8.60–13.80)	0.374[1]
White blood cell ($\times10^3/\mu L$)	7.80 (5.50–16.50)[b]	8.46 (4.17–22.70)[b]	10.20 (5.50–23.00)[a]	<0.001[1]
Platelet ($\times10^3/\mu L$)	197.50 (85.00–303.00)[b]	200.00 (118.00–408.00)	214.00 (126.00–500.00)[a]	0.022[1]
Triglyceride (mg/dl)	176.50 (57.00–285.00)[c]	200.00 (58.00–347.00)[b]	233.00 (117.00–384.00)[a]	<0.001[1]
Total cholesterol (mg/dl)	179.50 (148.00–305.00)[b]	210.00 (123.00–360.00)[a]	207.00 (120.00–474.00)[a]	<0.001[1]

Data are presented as median (min–max) for continuous variables and as percentages (%) for categorical variables. BMI: Body-mass index, GWG: Gestational weight gain, 1 Kruskal–Wallis test, 2 Chi-square test p<0.05 was considered statistically significant. Superscripts a>b>c indicate significance in post-hoc Dunnett's T3 test

Table 2. Comparison of perinatal outcomes by BMI groups

Variable	BMI normal (n=70)	BMI overweight (n=70)	BMI obese (n=70)	p-value
<37 weeks delivery, n (%)	8 (11.4%)[b]	10 (14.3%)[b]	25 (35.7%)[a]	<0.001[2]
<34 weeks delivery, n (%)	2 (2.9%)	1 (1.4%)	8 (11.4%)	0.016[2]
Gestational age (weeks)	38.0 (32.0-41.0)[b]	38.0 (33.0-41.0)[b]	37.0 (32.0-42.0)[a]	<0.001[1]
Birth weight (g)	3242.5 (1910.0-4200.0)	3365.0 (1745.0-4590.0)	3262.5 (1430.0-4320.0)	0.061[1]
Low birth weight (<2500 g)	4 (5.7%)	2 (2.9%)	7 (10.0%)	0.211[2]
Macrosomia (≥4000 gr)	5 (7.1%)	7 (10.0%)	12 (17.1%)	0.160[2]
NICU admission, n (%)	8 (11.4%)	4 (5.7%)	10 (14.3%)	0.241[2]
RDS	1 (1.4%)	1 (1.4%)	2 (2.9%)	
TTN	4 (5.7%)	2 (2.9%)	2 (2.9%)	
Hypoglycemia	1 (1.4%)	1 (1.4%)	2 (2.9%)	
Hyperbilirubinemia	2 (2.9%)	0 (0%)	3 (4.3%)	
Sepsis	0 (0%)	0 (0%)	1 (1.4%)	
Delivery type, n (%)				
Vaginal	47 (67.1%)[a]	40 (57.1%)	23 (32.9%)[b]	<0.001[2]
Cesarean	23 (32.9%)[b]	30 (42.9%)	47 (67.1%)[a]	<0.001[2]
Other obstetric complications, n (%)	5 (%7.1)	2 (2.9%)	8 (%11.9)	0.144[2]
PPROM	3 (2.9%)	1 (1.4%)	1 (1.4%)	
Preeclampsia	0 (0%)	0 (0%)	4 (5.7%)	
IUGR	0 (0%)	0 (0%)	2 (2.9%)	
Placenta previa	2 (2.9%)	0 (0%)	0 (0%)	
Placental abruption	0 (0%)	1 (1.4%)	0 (0%)	
IUFD	0 (0%)	0 (0%)	1 (1.4%)	
APGAR score at 1st minute	9.0 (7.0-9.0)	9.0 (7.0-9.0)	9.0 (0.0-9.0)	0.117[1]
APGAR score at 5th minute	10.0 (8.0-10.0)[b]	10.0 (8.0-10.0)[b]	10.0 (0.0-10.0)[a]	0.007[1]

Data are presented as median (min–max) for continuous variables and as percentages (%) for categorical variables. 1 Kruskal–Wallis test, 2 Chi-square test p<0.05 was considered statistically significant. Superscripts a>b indicate significance in post-hoc Dunnett's T3 test. Pairwise comparisons of significantly different categorical variables were performed using Bonferroni-adjusted Chi-square or Fisher's exact test as appropriate. Abbreviations: BMI: Body-mass index, NICU: Neonatal intensive care unit, RDS: Respiratory distress syndrome, TTN: Transient tachypnea of the newborn, PPROM: Preterm premature rupture of membranes, IUGR: Intrauterine growth restriction, IUFD: Intrauterine fetal demise, APGAR: Appearance, pulse, grimace, activity, respiration

was not statistically significant (p=0.160). NICU admission rates were higher in the obese group (14.3%) but did not reach statistical significance (p=0.241). The 1-minute Apgar score did not differ among the groups (p=0.117), while the 5-minute Apgar score was significantly lower in the obese group (p=0.007). Obstetric complications such as preeclampsia (5.7%), IUGR (2.9%), and IUFD (1.4%) were observed only in the obese group. Cesarean delivery was significantly more frequent in the obese group (p<0.001).

Additionally, Spearman correlation analysis was performed to assess the relationship between GWG and key perinatal outcomes. As shown in **Supplementary Table 1**, a weak positive correlation was observed between GWG and gestational age at birth (r=0.122, p=0.077), although it did not reach statistical significance. No significant correlations were found between GWG and birth weight, NICU admission, or Apgar scores at 1 and 5 minutes.

Univariable Regression Analysis (Model 1)

In univariable regression (**Table 3**), obesity was significantly associated with an increased risk of <37-week delivery (OR: 4.31, 95% CI: 1.78–10.42, p=0.001) and lower 5-minute Apgar score (β=-0.3, 95% CI: -0.6 to -0.0, p=0.025). Although not statistically significant, obesity was associated with higher odds of macrosomia (OR: 2.69, 95% CI: 0.89–8.09, p=0.078). No statistically significant associations were observed for <34-week delivery, low birth weight, NICU admission, or 1-minute Apgar score.

Multivariable Regression Analysis (Model 2)

After adjusting for maternal age, fasting glucose, smoking, treatment modality, and GWG (**Table 4**), the following results were obtained:

- The association between obesity and <37-week delivery remained significant, and the adjusted odds ratio further increased (OR:6.62, 95% CI:2.11–20.74, p=0.001).

- <34-week delivery and low birth weight (<2500 g), which were not significant in model 1, became statistically significant in model 2 (p=0.012 and p=0.026, respectively).

- The association between obesity and macrosomia was not significant after adjustment (aOR:1.12, 95% CI: 0.27–4.62, p=0.880).

- Associations with NICU admission, 1-minute, and 5-minute Apgar scores were not significant in the adjusted model.

Additionally, Spearman correlation analysis was performed to assess the relationship between GWG and key perinatal outcomes. As shown in **Supplementary Table 1**, a weak positive correlation was observed between GWG and gestational age at birth (r=0.122, p=0.077), although it did not reach statistical significance. No significant correlations were found between GWG and birth weight, NICU admission, or Apgar scores at 1 and 5 minutes.

Visual Summary

The forest plots presented in **Figure 1** compare the univariable (model 1) and multivariable (model 2) regression results for the

Supplementary Table 1. Correlation between gestational weight gain (GWG) and perinatal outcomes

Variable	Gestational age at birth (weeks) r (p)	Birth weight (g) r (p)	NICU admission r (p)	Apgar score (1 min) r (p)	Apgar score (5 min) r (p)
GWG	0.122 (0.0772)	-0.061 (0.3773)	0.025 (0.719)	0.012 (0.8611)	0.012 (0.8611)

Statistical test: Spearman correlation analysis. r, Spearman correlation coefficient; p, p-value. Abbreviations: GWG: Gestational weight gain, NICU: Neonatal intensive care unit, Apgar: Appearance, pulse, grimace, activity, and respiration

Table 3. Model 1–univariable logistic and linear regression results

Outcome	Comparison group	OR or beta	95% confidence interval	p-value
Preterm <37 wks	Overweight	1.29	0.48–3.49	0.614
Preterm <37 wks	Obese	4.31	1.78-10.42	0.001
Preterm <34 wks	Overweight	0.49	0.04 – 5.56	0.567
Preterm <34 wks	Obese	4.39	0.90-21.45	0.068
Birth weight <2500g	Overweight	0.49	0.09-2.74	0.413
Birth weight <2500g	Obese	1.83	0.51-6.57	0.352
Macrosomia (>4000 g)	Overweight	1.44	0.44-4.79	0.55
Macrosomia (>4000 g)	Obese	2.69	0.89-8.09	0.078
NICU admission	Overweight	0.47	0.13-1.64	0.236
NICU admission	Obese	1.29	0.48-3.49	0.614
Apgar 1 min	Overweight	0.0	-0.2-0.3	0.747
Apgar 1 min	Obese	-0.2	-0.5-0.0	0.086
Apgar 5 min	Overweight	0.0	-0.2-0.3	0.915
Apgar 5 min	Obese	-0.3	-0.6--0.0	0.025

Reference group: Normal weight, GDM; Model 1: OR values are derived from univariable logistic regression models. Apgar scores are continuous variables and were analyzed using linear regression. Reported values for Apgar scores represent β (beta) coefficients (mean difference). Abbreviations: NICU: Neonatal intensive care unit, APGAR: Appearance, pulse, grimace, activity, respiration

Table 4. Model 2–multivariable logistic and linear regression results

Outcome	Comparison group	OR or beta	95% confidence interval	p-value
Preterm <37 wks	Overweight	1.45	0.52-4.07	0.482
Preterm <37 wks	Obese	6.62	2.11-20.74	0.001
Preterm <34 wks	Overweight	0.66	0.06-7.75	0.741
Preterm <34 wks	Obese	12.00	1.74-82.63	0.012
Birth weight <2500g	Overweight	0.76	0.13-4.51	0.759
Birth weight <2500g	Obese	6.66	1.25-35.34	0.026
Macrosomia (>4000 g)	Overweight	1.10	0.32-3.83	0.880
Macrosomia (>4000 g)	Obese	1.12	0.27-4.62	0.880
NICU admission	Overweight	0.52	0.14-1.9	0.324
NICU admission	Obese	1.93	0.52-7.11	0.322
Apgar 1 min	Overweight	0.08	-0.13-0.28	0.446
Apgar 1 min	Obese	-0.04	-0.29-0.21	0.744
Apgar 5 min	Overweight	0.09	-0.1-0.28	0.362
Apgar 5 min	Obese	0.0	-0.23-0.24	0.975

Reference group: Normal weight GDM; Model 2: OR values are derived from multivariable logistic regression models adjusted for age, fasting glucose, smoking, treatment type, GWG. Apgar scores are continuous variables and were analyzed using linear regression. Reported values for Apgar scores represent β (beta) coefficients (mean difference). Abbreviations: NICU: Neonatal intensive care unit, APGAR: Appearance, pulse, grimace, activity, respiration

obese group. In the logistic regression panel (A), the adjusted odds ratios (aORs) for <37-week and <34-week deliveries, as well as low birth weight (<2500 g), were higher in model 2. Among these, <34-week delivery and low birth weight reached statistical significance only after adjustment. Regarding macrosomia, although a positive association was observed in model 1 (OR:2.69), it did not reach statistical significance and the association was further attenuated in model 2 (aOR:1.12). In the linear regression panel (B), the association between obesity and the 5-minute Apgar score observed in model 1 disappeared after adjustment for covariates.

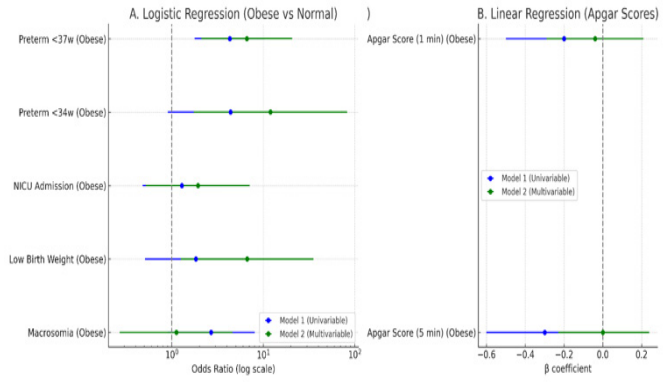


Figure 1. Forest plot showing the association between obesity and adverse perinatal outcomes in GDM patients
Forest plots illustrating the association between maternal obesity and adverse perinatal outcomes in pregnancies complicated by gestational diabetes mellitus (GDM). Panel A presents odds ratios (log scale) for obesity (vs. normal BMI) in relation to preterm birth (<37 and <34 weeks), NICU admission, low birth weight, and macrosomia, using both univariable (Model 1, blue) and multivariable (model 2, green) logistic regression analyses. Panel B shows β coefficients for 1-minute and 5-minute Apgar scores using linear regression, again comparing model 1 and model 2. Confidence intervals are displayed for all estimates.

To provide a visual summary of distribution patterns, a heatmap was created to display the frequency of selected perinatal outcomes across BMI categories (**Figure 2**). The highest rates of preterm delivery (<37 and <34 weeks), macrosomia (≥4000 g), NICU admission, and cesarean section were observed in the obese group. In contrast, the normal weight group showed the highest rates of vaginal delivery. This figure highlights the overall increase in adverse outcomes with increasing BMI.

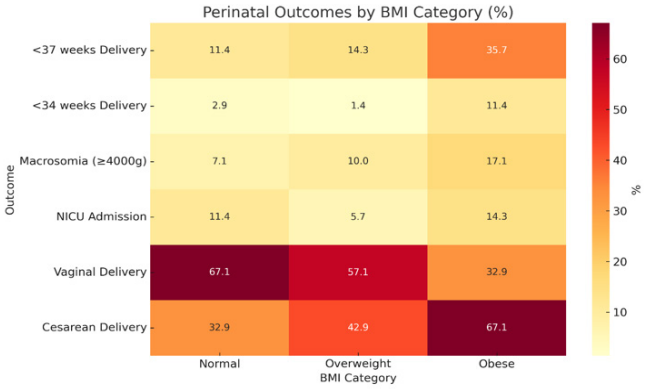


Figure 2. Heatmap visualization of selected perinatal outcomes across BMI categories
Heatmap visualization of selected perinatal outcomes across different BMI categories in women with gestational diabetes mellitus (GDM). The chart illustrates the percentage distribution of outcomes including <37 weeks and <34 weeks deliveries, macrosomia (≥4000g), NICU admission, vaginal delivery, and cesarean delivery across normal, overweight, and obese BMI groups. Darker shades represent higher percentages. Notably, the obese group demonstrated the highest rates of preterm delivery, macrosomia, NICU admission, and cesarean section.

DISCUSSION

Our study demonstrated that preterm birth rates before both 37 and 34 weeks of gestation were significantly higher among obese pregnant women with GDM. This finding suggests that maternal obesity may increase the risk of early delivery in pregnancies complicated by GDM. Increased systemic inflammation, insulin resistance, and endothelial dysfunction associated with obesity may trigger the preterm birth process by impairing uteroplacental perfusion and promoting early placental aging. As consistently noted in the literature, both early and late pregnancy complications are more frequently observed in obese pregnant women, and preterm birth stands out as one of the most significant outcomes.[16,17] This is in line with our findings. A recent large-scale prospective cohort study provided strong support for this association by showing that maternal obesity in combination with GDM significantly

increased the risk of both spontaneous (OR 1.98; 95% CI: 1.13–3.47) and medically indicated (OR 2.05; 95% CI: 1.25–3.37) preterm births.[18] These findings support the idea that the coexistence of obesity and GDM may act through multiple biological pathways, including inflammation, vascular dysfunction, and placental aging, to increase preterm birth risk. Conversely, one study reported that obesity was not significantly associated with spontaneous preterm birth in women with GDM but was linked only to indicated preterm birth.[19] Such discrepancy may be attributed to differences in sample characteristics, study design, and definitions of preterm birth. Nevertheless, it is widely recognized that the metabolic disturbances caused by obesity, particularly in the presence of GDM, may amplify inflammatory mechanisms and increase the likelihood of preterm delivery.

In our study, the rate of low birth weight (<2500 g) was significantly higher among obese women with GDM, while macrosomia (birthweight ≥4000 g) was more frequent in this group, albeit not reaching statistical significance. These findings suggest that maternal obesity may exert dual and heterogeneous effects on fetal growth. On the one hand, obesity-related metabolic dysfunctions may impair uteroplacental blood flow and nutrient transfer, restricting fetal growth. On the other hand, heightened insulin resistance and increased transplacental glucose transfer may predispose to excessive fetal growth. Several studies have reported that obese women had an increased risk of both intrauterine growth restriction (IUGR) and macrosomia.[20,21] These findings highlight the non-uniform nature of obesity's effects on fetal development and underscore the importance of considering individual metabolic profiles and placental function when interpreting fetal growth outcomes.

Our findings also revealed a significantly higher cesarean delivery rate among obese women with GDM compared to those with normal BMI. This result aligns with previous research indicating that maternal obesity is a strong independent risk factor for cesarean section.[22,23] Several mechanisms may contribute to this relationship, including increased incidence of labor dystocia, macrosomia, and a higher prevalence of pregnancy complications requiring surgical delivery. Furthermore, altered myometrial contractility and increased soft tissue impedance in obese women may reduce the likelihood of successful vaginal delivery. These factors collectively support the need for individualized intrapartum management strategies in obese pregnant women, particularly those with GDM.

In our study, NICU admission rates were numerically higher in the obese group compared to the normal BMI group, although this difference did not reach statistical significance. However, previous studies have demonstrated a clear association between maternal obesity and increased NICU admission risk.[24,25] For instance, a large retrospective analysis in Belgium showed that maternal obesity was independently associated with a 38% higher adjusted odds of NICU admission in neonates born to obese mothers.[24] Similarly, in the DEPOSIT cohort study, Ray et al.[25] reported that maternal obesity conferred a significantly higher risk of NICU admission, particularly in pregnancies complicated by diabetes. These findings underscore the need for diligent neonatal monitoring and preparedness in managing obese pregnancies complicated by GDM.

Additionally, we performed a Spearman correlation analysis to explore the association between GWG and key perinatal outcomes. The analysis revealed a weak positive correlation between GWG and gestational age at birth, which did not reach statistical significance ($r=0.122$, $p = 0.077$). However, no significant correlation was observed between GWG and birth weight, NICU admission, or Apgar scores. These results suggest that while GWG may have a modest influence on pregnancy duration, its impact on neonatal outcomes may be limited in this cohort. In contrast, Ke et al.[26] reported that excessive GWG was significantly associated with increased risks of macrosomia, LGA, and overall pregnancy complications among women with GDM, particularly when combined with obesity. Further studies with larger sample sizes are warranted to clarify the potential role of GWG in determining perinatal outcomes in pregnancies complicated by GDM.

In addition to medical interventions, lifestyle modifications play a crucial role in the management of maternal obesity, especially in pregnancies complicated by GDM. Evidence from a recent meta-analysis including over 40.000 pregnant women indicates that combined diet and physical activity interventions can reduce the incidence of GDM by approximately 18% compared with standard care (RR 0.82; 95% CI 0.74–0.94).[27] Preconception counseling, individualized dietary plans, and regular physical activity have also been shown to limit excessive GWG and reduce the risk of adverse perinatal outcomes. Therefore, integrating structured lifestyle interventions early in pregnancy may improve maternal-fetal outcomes and reduce the healthcare burden associated with obesity-related complications.

In recent years, novel non-invasive ultrasonographic techniques, such as fetal breathing movement (FBM) analysis and nasal flow Doppler, have been increasingly investigated for their potential to predict adverse perinatal outcomes, including preterm birth. A prospective multicenter cohort study demonstrated that combining the absence or irregularity of FBM with a nasal Doppler inspiration/expiration (I/E) ratio of <1.25 predicted preterm birth within 24 hours with 94.6% sensitivity.[28] These methods may provide additional insight into fetal well-being, particularly in high-risk pregnancies complicated by maternal obesity or GDM. Additionally, maternal nutritional quality assessed by validated scoring systems such as the healthy eating index (HEI) has also been linked to both fetal growth patterns and gestational age at delivery. For example, in a prospective multi-ethnic cohort study, lower HEI-2010 scores during pregnancy were associated with a 1.76-fold increased risk of delivering a large-for-gestational-age (LGA) infant. This finding highlights the role of maternal diet quality in modulating both fetal development and the timing of birth.

One of the strengths of our study is the use of first-trimester BMI measurements, which are less influenced by gestational weight changes and thus offer a more accurate assessment of pre-pregnancy obesity. Additionally, the study controlled

for important confounding variables such as maternal age, fasting glucose levels, HbA1c, smoking status, GWG, and treatment modality, which enhances the robustness of the results. Another strength is the homogeneous selection of patients with GDM according to IADPSG criteria and the stratified analysis of perinatal outcomes.

Limitations

However, our study has some limitations. First, its retrospective and single-center design may limit generalizability and introduce information bias. Second, we lacked detailed data on fetal well-being assessments such as biophysical profiles or fetal Doppler findings, which could have enriched the interpretation of neonatal outcomes. Third, although nutritional status likely plays a role in fetal development, we were unable to incorporate standardized dietary assessment tools such as the HEI due to data unavailability. Lastly, while our sample size was adequate for primary outcomes, it may not have been powered to detect subtle differences in some secondary outcomes like NICU admission or macrosomia.

CONCLUSION

This study highlights the significant impact of maternal obesity—defined by first-trimester BMI—on adverse perinatal outcomes in pregnancies complicated by GDM. Our findings indicate that obese women with GDM are at higher risk for preterm birth, low birth weight, and cesarean delivery. Although NICU admission and macrosomia rates were numerically higher in this group, these differences did not reach statistical significance. By using first-trimester BMI values, our study underscores the importance of early pregnancy weight assessment, and indirectly, the potential benefit of optimizing maternal weight even before conception. These results suggest that both preconceptional and early antenatal weight management strategies may help improve perinatal outcomes in women with GDM. Future prospective studies with larger cohorts are needed to validate these findings and guide clinical recommendations.

ETHICAL DECLARATIONS

Ethics Committee Approval

This study has been approved by the Scientific Researches Ethics Committee of Bursa City Hospital (Date: 14.05.2025, Decision No: 2025/10-12).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Wendland EM, Torloni MR, Falavigna M, et al. Gestational diabetes and pregnancy outcomes--a systematic review of the World Health Organization (WHO) and the International Association of Diabetes in Pregnancy Study Groups (IADPSG) diagnostic criteria. *BMC Pregnancy Childbirth*. 2012;12(1):23. doi:10.1186/1471-2393-12-23
2. Wang H, Li N, Chivese T, et al. IDF diabetes atlas: estimation of global and regional gestational diabetes mellitus prevalence for 2021 by international association of diabetes in pregnancy study group's criteria. *Diabetes Res Clin Pract*. 2022;183:109050. doi:10.1016/j.diabres.2021.109050
3. Gao C, Sun X, Lu L, et al. Prevalence of gestational diabetes mellitus in mainland China: a systematic review and meta- analysis. *J Diabetes Investig*. 2019;10(1):154-162. doi:10.1111/jdi.12854
4. Yen IW, Lee CN, Lin MW, et al. Overweight and obesity are associated with clustering of metabolic risk factors in early pregnancy and the risk of GDM. *PLoS One*. 2019;14(12):e0225978. doi:10.1371/journal.pone.0225978
5. Pantham P, Aye IL, Powell TL. Inflammation in maternal obesity and gestational diabetes mellitus. *Placenta*. 2015;36(7):709-715. doi:10.1016/j.placenta.2015.04.006
6. Šimják P, Cinkajzlová A, Anderlová K, et al. The role of obesity and adipose tissue dysfunction in gestational diabetes mellitus. *J Endocrinol*. 2018;238(2):R63-77. doi:10.1530/JOE-18-0032
7. Catalano PM, McIntyre HD, Cruickshank JK, et al. The hyperglycemia and adverse pregnancy outcome study: associations of GDM and obesity with pregnancy outcomes. *Diabetes Care*. 2012;35(4):780-786. doi:10.2337/dc11-1790
8. Wahabi HA, Fayed AA, Alzeidan RA, Mandil AA. The independent effects of maternal obesity and gestational diabetes on the pregnancy outcomes. *BMC Endocr Disord*. 2014;14(1):47. doi:10.1186/1472-6823-14-47
9. Martin KE, Grivell RM, Yelland LN, Dodd JM. The influence of maternal BMI and gestational diabetes on pregnancy outcome. *Diabetes Res Clin Pract*. 2015;108(3):508-513. doi:10.1016/j.diabres.2014.12.015
10. Ijäs H, Koivunen S, Raudaskoski T, Kajantie E, Gissler M, Vääräsmäki M. Independent and concomitant associations of gestational diabetes and maternal obesity to perinatal outcome: a register-based study. *PLoS One*. 2019;14(8):e0221549. doi:10.1371/journal.pone.0221549
11. Weschenfelder F, Hein F, Lehmann T, Schleußner E, Groten T. Clinical medicine contributing factors to perinatal outcome in pregnancies with gestational diabetes: what matters most? A retrospective analysis. *J Clin Med*. 2021;10(2):348. doi:10.3390/jcm10020348
12. Whiteman VE, Salemi JL, Mejia De Grubb MC, et al. Additive effects of pre-pregnancy body-mass index and gestational diabetes on health outcomes and costs. *Obesity (Silver Spring)*. 2015;23(11):2299-308. doi:10.1002/oby.21222
13. Hartling L, Dryden DM, Guthrie A, et al. Screening and diagnosing gestational diabetes mellitus. *Evid Rep Technol Assess (Full Rep)*. 2012;(210):1-327.
14. Kominiarek MA, Saade G, Mele L, et al. Association between gestational weight gain and perinatal outcomes. *Obstet Gynecol*. 2018;132(4):875-881. doi:10.1097/AOG.0000000000002854
15. Sun Y, Shen Z, Zhan Y, et al. Effects of pre-pregnancy body-mass index and gestational weight gain on maternal and infant compli cations. *BMC Pregnancy Childbirth*. 2020;20(1):390. doi:10.1186/s12884-020-03071-y
16. Zhu Y, Zheng Q, Pan Y, et al. Association between prepregnancy body mass index or gestational weight gain and adverse pregnancy outcomes among Chinese women with gestational diabetes mellitus: a systematic review and meta-analysis. *BMJ Open*. 2024;14(2):e075226. doi:10.1136/bmjopen-2023-075226
17. Su WJ, Chen YL, Huang PY, et al. Effects of prepregnancy body mass index, weight gain, and gestational diabetes mellitus on pregnancy outcomes: a population-based study in Xiamen, China, 2011-2018. *Ann Nutr Metab*. 2019;75(1):31-38. doi:10.1159/000501710

18. Liu K, Chen Y, Tong J, Yin A, Wu L, Niu J. Association of maternal obesity with preterm birth phenotype and mediation effects of gestational diabetes mellitus and preeclampsia: a prospective cohort study. *BMC Pregnancy Childbirth*. 2022;22(1):459. doi:10.1186/s12884-022-04780-2
19. Smith GC, Shah I, Pell JP, Crossley JA, Dobbie R. Maternal obesity in early pregnancy and risk of spontaneous and elective preterm deliveries: a retrospective cohort study. *Am J Public Health*. 2007;97(1):157-162. doi: 10.2105/AJPH.2005.074294
20. Lewandowska M, Sajdak S, Więckowska B, Manevska N, Lubiński J. The influence of maternal BMI on adverse pregnancy outcomes in older women. *Nutrients*. 2020;12(9):2838. doi:10.3390/nu12092838
21. Liu P, Xu L, Wang Y, et al. Association between perinatal outcomes and maternal pre-pregnancy body-mass index. *Obes Rev.* 2016;17(11):1091-1102. doi:10.1111/obr.12455
22. Kominiarek MA, Vanveldhuisen P, Hibbard J, et al. Consortium on safe labor. The maternal body mass index: a strong association with delivery route. *Am J Obstet Gynecol*. 2010;203(3):264.e1-267. doi:10.1016/j.ajog.2010.06.024
23. Sebire NJ, Jolly M, Harris JP, et al. Maternal obesity and pregnancy outcome: a study of 287,213 pregnancies in London. *Int J Obes Relat Metab Disord*. 2001;25(8):1175-1182. doi:10.1038/sj.ijo.0801670
24. Bogaerts A, Van den Bergh B, Nuyts E, Martens E, Witters I, Devlieger R. Socio-demographic and obstetrical correlates of pre-pregnancy body-mass index and gestational weight gain. *Clin Obes*. 2012;2(5-6):150-159. doi:10.1111/cob.12004
25. Ray JG, Vermeulen MJ, Shapiro JL, Kenshole AB. Maternal and neonatal outcomes in pregestational and gestational diabetes mellitus, and the influence of maternal obesity and weight gain: the DEPOSIT study. Diabetes endocrine pregnancy outcome study in Toronto. *QJM*. 2001; 94(7):347-356. doi:10.1093/qjmed/94.7.347
26. Ke JF, Liu S, Ge RL, Ma L, Li MF. Associations of maternal pre-pregnancy BMI and gestational weight gain with the risks of adverse pregnancy outcomes in Chinese women with gestational diabetes mellitus. *BMC Pregnancy Childbirth*. 2023;23(1):414. doi:10.1186/s12884-023-05657-8
27. Takele WW, Vesco KK, Josefson J, et al. Effective interventions in preventing gestational diabetes mellitus: a systematic review and meta-analysis. *Commun Med (Lond).* 2024;4(1):75. doi:10.1038/s43856-024-00491-1
28. Esin S, Okuyan E, Gunakan E, Zengin HY, Hayran M, Tohma YA. A novel technique for prediction of preterm birth: fetal nasal flow Doppler. *J Perinat Med*. 2020;49(3):319-325. doi:10.1515/jpm-2020-0276
29. Zhu Y, Hedderson MM, Sridhar S, Xu F, Feng J, Ferrara A. Poor diet quality in pregnancy is associated with increased risk of excess fetal growth: a prospective multi-racial/ethnic cohort study. *Int J Epidemiol*. 2019;48(2):423-432. doi:10.1093/ije/dyy285

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1715880

Anatolian Curr Med J. 2025;7(5):559-562

Impact of JAK2 V617F allele burden on clinical and laboratory parameters in polycythemia vera

Özlem Beyler, Cengiz Demir

Department of Hematology, Gazi Yaşargil Training and Research Hospital, Diyarbakır, Turkiye

Cite this article as: Beyler Ö, Demir C. Impact of JAK2 V617F allele burden on clinical and laboratory parameters in polycythemia vera. *Anatolian Curr Med J.* 2025;7(5):559-562.

Received: 08.06.2025 ◆ **Accepted**: 24..07.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: To investigate the association between the JAK2 V617F allele burden and clinical and laboratory parameters in polycythemia vera (PV) patients, focusing on its association with thrombotic events, disease severity, and systemic inflammatory markers.

Methods: This retrospective study included 71 patients with PV. Data were collected from medical records, including demographics, laboratory values, spleen size, and thrombotic history. Patients were stratified by JAK2 V617F allele burden into subgroups for comparative analysis. Quantitative polymerase chain reaction (PCR) was used to measure allele burden. Statistical analyses were performed using the Mann-Whitney U test, Kruskal-Wallis test, and Spearman correlation, with a significance level of $p<0.05$.

Results: The median JAK2 V617F allele burden was significantly higher in women than in men (9.53% vs. 2.00%, p=0.003). Patients with platelet counts ≥400×10^9/L had a significantly higher allele burden than those with counts <400×10^9/L (8.94% vs. 1.33%, p=0.019). There was no significant association between allele burden and thrombotic events (p=0.549) or splenomegaly (p=0.191).

Conclusion: JAK2 V617F allele burden is associated with certain laboratory parameters, including platelet count, and varies by gender in patients with PV. Although not significantly associated with thrombotic events or splenomegaly in this cohort, allele burden remains a potentially valuable biomarker for disease monitoring and individualised treatment.

Keywords: Polycythemia vera, janus kinase 2, alleles

INTRODUCTION

Polycythemia vera (PV) is the most common subtype among myeloproliferative neoplasms, characterized by the clonal proliferation of hematopoietic stem cells. The disease presents with an increase in red blood cell mass, leukocytosis, thrombocytosis, and an increased risk of thrombotic and hemorrhagic complications. The diagnosis is supported by findings such as increased blood cell production, the presence of JAK2 mutations, and bone marrow biopsy results (**Table 1**).[1,2] The pathogenesis and natural course of PV have been better elucidated through understanding the genetic and biological factors associated with disease progression.

PV is driven by constitutive activation of the JAK-STAT signaling pathway due to the JAK2 V617F mutation.[1,2] This pathway activation promotes clonal hematopoiesis and inflammatory cytokine production, contributing not only to myeloproliferation but also to disease related complications. Over time, chronic myeloproliferation may lead to stem cell exhaustion, clonal evolution, and epigenetic alterations, increasing the risk of post-polycythemic myelofibrosis and, less frequently, transformation to acute myeloid leukemia.[1,2]

Table 1. Diagnostic criteria for polycythemia vera

Criterion type	Criteria	Diagnostic thresholds
Major criteria 1	Elevated hemoglobin, hematocrit, or red cell mass	- Hemoglobin: >16.5 g/dl (male), >16.0 g/dl (female) - Hematocrit: >49% (male), >48% (female) - Red cell mass: >25% above the predicted normal value
Major criteria 2	Presence of JAK2 mutation	JAK2 V617F or JAK2 exon 12 mutation
Major criteria 3	Hypercellular bone marrow biopsy showing trilineage proliferation (erythroid, granulocytic, and megakaryocytic) without atypia	Not required in cases with hemoglobin >18.5 g/dl (male) or >16.5 g/dl (female) and hematocrit >55.5% (male) or >49.5% (female) with JAK2 mutation
Minor criteria	Serum erythropoietin level	Below the normal range
Diagnostic requirements	- All major criteria must be met, or - The first two major criteria and the minor criterion must be met	

The JAK2 V617F mutation, is detected in over 95% of cases and serves as a critical biomarker in diagnosis.[1-3] JAK2 V617F

Corresponding Author: Özlem Beyler, drozlembeyler@gmail.com

allele burden has been linked with both clinical features and prognosis of the disease.[3]

This study aims to investigate the relationship between JAK2 V617F allele burden and clinical and laboratory parameters in PV in greater detail. By focusing on the correlation of allele burden with thrombotic events, disease severity, and systemic inflammatory markers, the study seeks to evaluate the potential role of this biomarker in disease management and prognosis.

METHODS

This study included 71 patients diagnosed with PV. The diagnostic and clinical data of the patients were retrospectively collected from medical records. The study was approved by the Clinical Researches Ethics Committee of Gazi Yaşargil Training and Research Hospital. (Date: 27.12.2024, Decision No: 283), and informed consent was obtained from all participants. All procedures were followed with ethical guidelines and the tenets of the Declaration of Helsinki. Patients with JAK2 exon 12 mutations were excluded from the study. The diagnosis of PV was made according to the most recent WHO criteria (**Table 1**). At diagnosis, white blood cell count (WBC), neutrophil, lymphocyte, monocyte, hemoglobin (Hb), and platelet (PLT) counts and serum erythropoietin levels were recorded. Spleen size was assessed by abdominal ultrasonography, with lengths over 120 mm considered splenomegaly. A detailed history of thrombosis, including arterial and venous events, was obtained.

The JAK2 V617F mutation was detected and quantified using a Quantitative polymerase chain reaction (PCR)-based method. Genomic DNA was extracted from peripheral blood and PCR amplification was performed using specific primers. Mutation burden was calculated by comparing mutant and total allele products via a standard curve.

Statistical Analysis

The data analyses were conducted using SPSS Version 25.0. Categorical variables were summarized as frequencies (%), and continuous variables as medians (min–max). The Mann–Whitney U and Kruskal–Wallis tests were used for group comparisons. Correlations were assessed using Spearman's rho, and categorical comparisons with Pearson's Chi-square or Fisher's exact test. A p-value <0.05 was considered statistically significant.

RESULTS

The study included 71 patients with a mean age of 58 years (range: 20–90). Of these, 32 (45.1%) were female, and 39 (54.9%) were male. The median JAK2 V617F allele burden among all patients was 14.42%, which was used as the cutoff value for subgroup comparisons. The median JAK2V617F allele burden was 9.53% (range: 0.5–54%) in females and 2.00% (range: 0.01–25%) in males, a statistically significant difference (p=0.003).

Patients were grouped by age: 46 patients (64.8%) were ≤60 years old, and 25 patients (35.2%) were >60 years old. The median JAK2V617F allele burden in patients >60 years of age was 9.98% (range: 0.01–22.69%), compared to 4.96% (range: 0.01–54.00%) in those ≤60 years of age, a statistically significant difference (p=0.039). Gender distribution did not differ significantly between age groups.

Based on Hb levels, patients were categorized into two groups: 9 patients (12.7%) with Hb <17 g/dl and 62 patients (87.3%) with Hb≥17 g/dl. The median JAK2V617F allele burden was 5.55% (range: 0.50–22%) in patients with Hb <17 g/dl and 8.39% (range: 0.01–54%) in those with Hb≥17 g/dl (p=0.653).

The median WBC count for all patients was 11×10^9/L (range: 5.8–21×10^9/L). Patients were divided into two groups: 37 patients (52.1%) with WBC<10×10^9/L and 34 patients (47.9%) with WBC≥10×10^9/L. The median JAK2V617F allele burden was 5.60% (range: 0.01-33.40%) in patients with WBC<10×10^9/L and 8.75% (range: 0.01–54%) in those with WBC ≥10×10^9/L (p=0.180).

The median PLT count for all patients was 528×10^9/L (range: 161–1.134×10^9/L). Patients were classified into two groups: 26 patients (36.6%) with PLT <400×10^9/L and 45 patients (63.4%) with PLT ≥400×10^9/L. The median JAK2V617F allele burden was 1.33% (range: 0.01–25%) in patients with PLT <400×10^9/L and 8.94% (range: 0.03–54%) in those with PLT ≥400×10^9/L, a statistically significant difference (p=0.019).

Spleen size was assessed in all patients. Splenomegaly was absent in 55 patients (77.5%) and present in 16 patients (22.5%). The median JAK2V617F allele burden was 9% (range: 0.05–20.25%) in patients with splenomegaly and 4.48% (range: 0.01–54%) in those without splenomegaly (p=0.191). Male patients had significantly larger spleens compared to females (p=0.003), with a mean difference of approximately 4.66 mm.

During the course of the disease, 15 patients (21.1%) had a history of thrombosis, while 56 patients (78.9%) did not. The median JAK2V617F allele burden was 9.09% (range: 0.01–33.4%) in patients with a history of thrombosis and 6.70% (range: 0.01–54%) in those without (p=0.549). Comparison of clinical and laboratory parameters by median JAK2 allele burden is shown in **Table 2**.

DISCUSSION

The clinical course varies of PV due to genetic and phenotypic heterogeneity. While some patients have a stable disease course, others are at risk of serious complications such as thrombosis, cardiovascular events, myelofibrosis or transformation into acute leukaemia. This variability necessitates personalised disease management and treatment strategies.

In recent years, JAK2 V617F allele burden has emerged as a critical biomarker for understanding the impact of PV on laboratory and clinical parameters. Higher allele burden has been associated with an increased risk of thrombosis, splenomegaly, myeloproliferation, and poor prognosis.[4-6,10] However, the mechanisms underlying these associations and their implications for clinical management remain incompletely understood. A better understanding of the relationship between JAK2 allele burden and clinical and laboratory parameters is essential to improve risk stratification and optimise treatment strategies.

Table 2. Comparison of clinical and laboratory parameters by median JAK2 allele burden

Variables	JAK2 median (min-max)	p
Gender, n (%)		
Female	9.53 (0.50-54.00)	0.003
Male	2.00 (0.01-25.00)	
Age, n (%)		
≤60	4.96 (0.01-54.00)	0.039
>60	9.98 (0.01-22.69)	
Hb, n (%)		
<16	5.55 (0.50-22.00)	0.653
≥16	8.39 (0.01-54.00)	
WBC, n (%)		
<10	5.60 (0.01-33.40)	0.180
≥10	8.75 (0.01-54.00)	
PLT, n (%)		
<400	1.33 (0.01-25.00)	0.019
≥400	8.94 (0.03-54.00)	
Epo, n (%)		
<5	8.58 (0.01-54.00)	0.135
≥5	1.92 (0.01-33.40)	
Spleen, n (%)		
≤120	4.48 (0.01-54.00)	0.191
>120	9.00 (0.05-20.25)	
Thrombosis, n (%)		
Absent	6.70 (0.01-54.00)	0.549
Present	9.09 (0.01-33.40)	

Mann-Whitney U test; p<0.05 is considered statistically significant. Min: Minimum, Max: Maximum, Hb: Hemoglobin, WBC: White blood cell count, PLT: Platelet

Studies have demonstrated correlations between JAK2 V617F allele burden and Hb, WBC, and PLT levels. Higher allele burden is generally associated with higher Hb and WBC levels, although the relationship with PLT counts varies across studies. In our study, the median allele burden was 14.42% and we observed a trend suggesting higher Hb levels might be associated with increased allele burden, although the result did not reach statistical significance. Similarly, Vannucchi et al.[4] reported a positive correlation between allele burden and Hb levels. Guglielmelli et al.[5] also found significantly higher Hb levels in patients with increased mutation burden.

Regarding WBC counts, Larsen[6] and Guglielmelli[5] found statistically significant differences in allele burden across different WBC levels. In our study, patients with WBC<10×10^9/L had a lower median JAK2 V617F allele burden compared to those with WBC ≥10×10^9/L, although the difference was not statistically significant. This lack of significance may be attributed to the relatively small sample size in our cohort.

In the analysis of PLT levels, we found that patients with PLT levels <400×10^9/L had significantly lower JAK2 V617F allele burden compared to those with PLT levels ≥400×10^9/L. This finding contrasts with studies by Larsen[6] and Guglielmelli,[5] who reported lower PLT counts in patients with higher allele burdens. Differences in patient populations and disease duration may account for these discrepancies.

Stein et al.[7] investigated the relationship between JAK2 V617F allele burden and gender in chronic myeloproliferative neoplasms, identifying gender as an independent factor influencing allele burden variability. Females were reported to have lower allele burdens than men, potentially due to less frequent mitotic recombination events in females. Similarly, Karantanos et al.[8] and Larsen et al.[6] found that women had lower allele burdens than males, suggesting sex-specific differences in disease biology. Interestingly, in our study, women exhibited significantly higher median JAK2 V617F allele burdens than men (p=0.003). This discrepancy highlights the variability of results between cohorts and underlines the need for careful interpretation. Factors such as sample size, genetic background, and disease duration may contribute to these differences.

Splenomegaly, present in approximately 36% of PV patients at diagnosis, is a significant clinical feature that influences prognosis and treatment decisions.[9] Vannucchi et al.[4] demonstrated a strong association between higher JAK2 V617F allele burden and the presence of splenomegaly, particularly in patients with allele burden >50%. Similarly, Guglielmelli et al.[5] found that palpable splenomegaly was more common in patients with allele burden >50%. In our study, patients with splenomegaly had a higher median allele burden (9.00%) compared to those without splenomegaly (4.48%), although the difference was not statistically significant (p=0.191).

Cardiovascular and thromboembolic events are leading causes of morbidity and mortality in PV. Current treatment strategies aim to prevent thrombotic complications through phlebotomy, low-dose aspirin, and cytoreductive therapy for high-risk patients.[1] Despite adherence to these treatments, thrombotic risk persists. Recent studies have focused on the impact of JAK2 V617F allele burden on thrombosis risk. Guglielmelli et al.[5] reported a higher incidence of venous thrombosis in patients with allele burden >50% compared to those with ≤50%. Similarly, Vannucchi et al.[4] found that an allele burden >75% increased thrombotic event risk. In our study, patients with thrombosis had higher allele burden than those without, although the difference was not statistically significant (p=0.549). The small sample size and the low proportion of patients with a high allele burden may explain this finding.

Limitations

Several limitations should be considered when interpreting the findings of this study. First, the retrospective design may introduce selection bias. Second, the relatively small sample size, particularly the small number of patients with high JAK2 allele load, may have reduced the statistical power to detect meaningful associations. Additionally, the lack of multivariate analysis prevented the adjustment for potential confounding factors such as age, gender, or treatment status. Finally, the absence of long-term data limits the ability to draw conclusions about prognostic value.

CONCLUSION

This study highlights the importance of JAK2 V617F allele burden in shaping the phenotypic variability of PV. The results suggest that allele burden may serve as a useful biomarker for predicting disease progression. However, larger cohorts and further research are needed to clarify the relationships between allele burden and clinical outcomes. Future studies should also include gender-specific analyses to improve our understanding of this complex disease.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was approved by the Clinical Researches Ethics Committee of Gazi Yaşargil Training and Research Hospital. (Date: 27.12.2024, Decision No: 283).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Spivak JL. How I treat polycythemia vera. *Blood*. 2019;134(4):341-352. doi:10.1182/blood.2018834044
2. Tefferi A, Vannucchi AM, Barbui T. Polycythemia vera: historical oversights, diagnostic details, and therapeutic views. *Leukemia*. 2021; 35(12):3339-3351. doi:10.1038/s41375-021-01401-3
3. Patel AB, Masarova L, Mesa RA, Hobbs G, Pemmaraju N. Polycythemia vera: past, present and future. *Leuk Lymphoma*. 2024;65(11):1552-1564. doi:10.1080/10428194.2024.2361836
4. Vannucchi AM, Antonioli E, Guglielmelli P, et al. Prospective identification of high-risk polycythemia vera patients based on JAK2(V617F) allele burden. *Leukemia*. 2007;21(9):1952-1959. doi:10.1038/sj.leu.2404854
5. Guglielmelli P, Loscocco GG, Mannarelli C, et al. JAK2V617F variant allele frequency >50% identifies patients with polycythemia vera at high risk for venous thrombosis. *Blood Cancer J*. 2021;11(12):199. doi:10.1038/s41408-021-00581-6
6. Larsen TS, Pallisgaard N, Møller MB, Hasselbalch HC. The JAK2 V617F allele burden in essential thrombocythemia, polycythemia vera and primary myelofibrosis--impact on disease phenotype. *Eur J Haematol*. 2007;79(6):508-515. doi:10.1111/j.1600-0609.2007.00960.x
7. Stein BL, Williams DM, Wang NY, et al. Sex differences in the JAK2 V617F allele burden in chronic myeloproliferative disorders. *Haematologica*. 2010;95(7):1090-1097. doi:10.3324/haematol.2009.014407
8. Karantanos T, Chaturvedi S, Braunstein EM, et al. Sex determines the presentation and outcomes in MPN and is related to sex-specific differences in the mutational burden. *Blood Adv*. 2020;4(12):2567-76. doi:10.1182/bloodadvances.2019001407
9. Tefferi A, Rumi E, Finazzi G, et al. Survival and prognosis among 1545 patients with contemporary polycythemia vera: an international study. *Leukemia*. 2013;27(9):1874-1881. doi:10.1038/leu.2013.163
10. Zhang Y, Zhou Y, Wang Y, et al. Thrombosis among 1537 patients with $JAK2^{V617F}$-mutated myeloproliferative neoplasms: risk factors and development of a predictive model. *Cancer Med*. 2020;9(6):2096-2105. doi:10.1002/cam4.2886

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1729371

Anatolian Curr Med J. 2025;7(5):563-568

Evaluating the information quality and readability on Turkish Websites about human papilloma virus

Cem Dağdelen[1], İlyas Turan[2], Özgür Ozan Ceylan[3], Evrim Erdemoğlu[1]

[1]*Department of Obstetrics and Gynecology, Faculty of Medicine, Süleyman Demirel University, Isparta, Turkiye*

[2] *Department of Gynecologic Oncology, Batman Training and Research Hospital, Batman, Turkiye*

[3]*Department of Obstetrics and Gynecology, Turgutlu State Hospital, Manisa, Turkiye*

Cite this article as: Dağdelen C, Turan İ, Ceylan ÖO, Erdemoğlu E. Evaluating the information quality and readability on Turkish Websites about human papilloma virus. *Anatolian Curr Med J.* 2025;7(5):563-568.

Received: 28.06.2025 • **Accepted**: 24..07.2025 • **Published**: 15.09.2025

ABSTRACT

Aims: We aimed to determine the quality and readability levels of the texts that health readers access from websites related to HPV.

Methods: 64 websites related to "human papilloma virus" were evaluated by using the Google search engine, in terms of readability and website information quality. The websites were divided into 6 groups according to their origins. Ateşman and Bezirci-Yılmaz readability formulas were used for readability, DISCERN measurement tool and JAMA criteria were used for website information quality and reliability. The information quality and reliability of the websites were evaluated by an obstetrician and gynecologist and a gynecological oncology surgeon.

Results: When the websites were evaluated according to the Ateşman readability formula, the readability score was found to be 57.2 and they were found to be moderately readable. According to the Bezirci-Yılmaz readability formula, the average readability score of all websites was found to be 12.8, that is, a readability at the undergraduate level was determined. When evaluated with the DISCERN tool, the average quality level of all websites was determined as "poor". It was seen that the texts prepared by the obstetrics and gynecology associations and the obstetricians and gynecologists were of higher quality than the other groups. It was observed that the quality of the texts was lower according to the gynecological oncology surgeon.

Conclusion: It is noteworthy that the texts prepared by the obstetrics and gynecology associations received higher quality scores than the other groups. Website information resources prepared by obstetrics and gynecology associations should be increased and the readability and quality of other internet health information needs to be improved.

Keywords: Gynecology, health readers, human papilloma virus

INTRODUCTION

The internet is increasingly used as a source of medical information both by healthcare professionals and patients due to its global accessibility, speed, and cost effectiveness.[1,2] While the rate of internet usage in Turkiye was 82.6% in the 16–74 age group in 2021, one of the most searched about topic on the internet was health-related information at a rate of 65.9%.[3,4]

In the United States, more than 61% of patients already use the Internet to inform themselves about their disease and possible treatment options.[5] Access to useful and understandable health information is an important factor when making health decisions. The internet is crucial to the modern dissemination of health information, but it is clear that the quality varies considerably between sources.[6] Numerous quality assessment tools have been developed to assess the quality of these resources.[7] In addition, written texts about health can cause patients to be misinformed, and lead to incorrect treatment and diagnosis.[8] In these cases, the readability of these texts, as well as their quality, is important.

The human papilloma virus (HPV) is the most common sexually transmitted infection.[9] In general, HPV is responsible for 90% of anal and cervical cancers and about 70% of vaginal and vulvar cancers. More than 200 HPV types have been identified and at least 14 high-risk types that can cause cancer were characterized. Two HPV types (16 and 18) cause approximately 70% of cervical cancers and precancerous cervical lesions.[10] It was determined that seeking information about HPV and sexually transmitted diseases from alternative sources, such as the internet, increases the anxiety of women.[11]

Colposcopic evaluation of malignant and premalignant epithelium of the vulva, vagina and cervix is performed according to certain visual features in terms of contour, color and vascular pattern using magnification after application of acetic acid or diluted Lugol's or Schiller's solution.[12]

Corresponding Author: İlyas Turan, ilyasturan04@gmail.com

In this study, it was aimed to determine the quality and readability levels of the texts that health readers access from websites related to HPV.

METHODS

Ethics

The study has been approved by the Clinical Researches Ethics Committee of Süleyman Demirel University (Date: 07.07.2022, Decision No: 190). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Included in this study were the first 100 websites that came up when typing the keywords "HPV" into the Google search engine on June 1st. Advertising sites, sites with no information about HPV, forum sites and scientific publications were excluded, and the remaining 64 websites were evaluated for readability and information quality. The origins of the websites (doctor, association, health institutions, newspaper, private laboratory, women's health websites) were recorded. The readability level of these recorded websites was determined and text quality analyses were performed.

While determining the readability level, the formulas of Ateşman and Bezirci-Yılmaz, which are frequently used in Turkish, were employed.

The information quality and reliability of the websites were evaluated by an obstetrician and gynecologist, and gynecological oncology surgeon. These evaluations were made by the two researchers independently, using the DISCERN measurement tool and Journal of the American Medical Association (JAMA) criteria.

Readability Formulas

The Ateşman (1997) readability formula was created by considering two variables, which were the sentence length and word length. The readability score is calculated using these variables.

The formula used to calculate the Ateşman readability score is as follows:

Readability score=198.825–40.175×(total syllables/total words)–2.610×(total words/total sentences).[13]

In the Bezirci-Yılmaz readability formula, variables such as the average sentence lengths and number of syllables affect the readability. The formula was created by using the number of words in the sentence and the number of syllables in the words. The readability formula is as follows:

Readability score=√OKS×((H3×0.84)+(H4×1.5)+(H5×3.5)+(H6×26.25))

Here, OKS is the average word count, H3 is the mean number of 3-syllable words, H4 is the mean number of 4-syllable words, H5 is the mean number of 5-syllable words, and H6 is the mean number of words with 6 or more syllables.[14]

The Bezirci-Yılmaz formula explains the readability level of a text by coinciding with a certain education level according to the education system in Turkiye (**Table 1**).

Table 1. Point of scores

Readability score	Readability level
90–100	Very easy
70–89	Easy
50–69	Moderate
30–49	Hard
1–29	Too hard

Quality Analysis of the Texts

Evaluation was made according to 4 criteria, including authorship, attribution, disclosure, and currency,which are included in the widely used JAMA criteria.[15] During the evaluation, scoring was done by giving 1 or 0 points for each criterion, depending on whether they met the criteria or not. According to these criteria, the score that a website can get as a result of evaluation is between 0 and 4 points.

The DISCERN questionnaire is a valid and reliable tool for analyzing health-related texts.[16] It can be used to evaluate the credibility of the website and whether the sources of evidence are clear. The DISCERN questionnaire, which consists of 3 sections, consists of 15 questions and 1 general evaluation. Questions 1 to 8 in section 1 evaluate the reliability of the website, questions 9 to 15 in section 2 are used to evaluate information about treatment options.[16] Section 3 measures the overall quality of the website. Each question is rated on a 5-point scale. An evaluation is made by giving 5 points if the text completely meets the criteria in question; 2, 3, or 4 points are given if it partially meets the criteria, and 1 point if the quality criterion is not met at all. The total DISCERN score is calculated by adding the ratings of the first 15 questions.18 The score can be categorized as excellent (63–75), good (51–62), moderate (39–50), low (27–38) or very low (<27).[17]

Statistical Analysis

The data obtained as a result of the study were analyzed with the PAWS Statistics 18 program. As a result of the analysis, the categorical variables were presented as the percentage and frequency, and the continuous variables were presented as the mean, standard deviation, median, minimum, and maximum. According to the Shapiro-Wilk test, the continuous variables did not show normal distribution ($p>0.05$). The Wilcoxon signed-rank test was used to compare the dependent variables, and the Kruskal-Wallis H test was used to compare more than two independent groups. The Mann-Whitney U test with Bonferroni correction was used for comparisons in groups with statistically significant differences as a result of the Kruskal-Wallis H test. Statistical significance was accepted as $p<0.05$.

RESULTS

In total, 64 websites were examined for readability and the quality of the texts. These 64 websites were evaluated by basically dividing them into 6 different groups according to their origins (**Table 2**). These groups consisted of obstetrics and gynecology associations, newspapers, websites related to women's health, private health institutions, obstetricians and

gynecologists, and private laboratories. Among the websites, the HPV-related texts were mostly created by obstetricians and gynecologists (Table 3).

Table 2. Education level

Grade	Education level
1–8	Primary education
9–12	Secondary education (high school)
12–16	Undergraduate
16+	Academic level

Table 3. Sources percentage of websites

	n	%
Obstetricians and gynecologists	21	32.8
Private health institution	17	26.6
Women's health websites	14	21.8
Private laboratories	6	9.4
Obstetrics and gynecology associations	3	4.7
Newspaper	3	4.7
Total	64	100.0

The websites were evaluated according to the readability formulas of Ateşman and Bezirci-Yılmaz. When the websites were evaluated according to the Ateşman readability formula, the readability score was 57.2, and they were moderately readable (Table 3). When the mean values between the groups were examined, it was observed that the texts on the websites created by obstetricians and gynecologists were the most easily readable, but when all of the groups were compared, it was seen that there was no difference in terms of readability (p=0.624) (Table 3).

According to the Bezirci-Yılmaz readability formula, the average readability score of all of the websites was 12.8, that is, a readability at the undergraduate level was determined. According to the Bezirci-Yılmaz readability formula, there was no difference in readability levels between the groups (p=0.534) (Table 3).

When the websites were evaluated with the JAMA criteria, according to both researchers, there was a statistically significant difference between the quality levels of the websites. According to both researchers, it was seen that the texts created by the newspapers had the lowest score, and the quality level of the texts created by the obstetrics and gynecology associations was the highest (Table 4). In the group evaluations, there was a statistically significant difference between the obstetrics and gynecology association and private health institution websites, between the obstetrician and gynecologist websites and the women's health websites and private health institution websites ($p<0.05$). There was a statistically significant difference between the obstetrics and gynecology association websites and the private health institution websites in the group comparisons according to the gynecological oncology surgeon ($p<0.05$).

When all of the websites were evaluated with the DISCERN tool, it was observed that they got 36 points according to the obstetrician and gynecologist, and 33 points when evaluated by the gynecological oncology surgeon. According to these results, the average quality level of all of the sites was determined as Poor. According to both researchers, it was seen that the texts prepared by the obstetrics and gynecology associations and obstetricians and gynecologists were of higher quality than the other groups. When evaluated by a gynecological oncology surgeon, it was determined that the average quality of the texts created by the newspapers was very poor (Table 5). In the group comparison, there was a statistically significant difference between the obstetrician and gynecologist websites and newspaper websites, according to both researchers.

Significant differences were found between the quality evaluations made by gynecological oncology surgeon and

Table 4. Formulation parameters of websites sources

		Mean	SD	Minimum	Maximum	p-value
Ateşman readability formula	Obstetrics and gynecology associations	53.8	5.3	50.3	60.0	0.624
	Obstetricians and gynecologists	59.4	8.7	41.3	76.8	
	Newspapers	57.7	5.4	51.5	61.0	
	Women's health	53.6	10.9	30.9	73.3	
	Private laboratories	56.8	11.1	44.5	72.1	
	Private health institutions	57.9	7.3	44.8	71.7	
	Total	57.2	8.9	30.9	76.8	
Bezirci-Yılmaz readability formula	Obstetrics and gynecology associations	13.7	0.5	13.2	14.3	0.534
	Obstetricians and gynecologists	12.2	2.7	7.1	17.2	
	Newspapers	11.4	2.0	10.0	13.7	
	Women's health	13.9	3.6	8.7	21.4	
	Private laboratories	13.6	3.4	9.2	17.5	
	Private health institutions	12.4	2.3	8.3	16.3	
	Total	12.8	2.8	7.1	21.4	

SD: Standard deviation

Table 5. Parameters of the type of specialist

		Mean	SD	Median	Minimum	Maximum	p-value
Obstetrician and gynecologist	Obstetrics and gynecology associations	2.0	0.0	2.0	2.0	2.0	<0.001
	Obstetricians and gynecologists	1.9	0.5	2.0	1.0	3.0	
	Newspapers	0.7	0.6	1.0	0.0	1.0	
	Women's health	1.6	1.1	2.0	0.0	3.0	
	Private laboratories	0.8	0.4	1.0	0.0	1.0	
	Private health institutions	1.0	0.5	1.0	0.0	2.0	
	Total	1.5	0.8	1.0	0.0	3.0	
Gynecological oncology surgeon	Obstetrics and gynecology associations	2.0	0.0	2.0	2.0	2.0	0.005
	Obstetricians and gynecologists	1.5	0.7	2.0	0.0	3.0	
	Newspapers	0.7	0.6	1.0	0.0	1.0	
	Women's health	1.6	1.1	2.0	0.0	3.0	
	Private laboratories	0.8	0.4	1.0	0.0	1.0	
	Private health institutions	0.8	0.5	1.0	0.0	2.0	
	Total	1.3	0.8	1.0	0.0	3.0	

SD: Standard deviation

obstetrician and gynecologist with JAMA (p=0.001) and DISCERN tool (p<0.001). It was observed that according to the gynecological oncology surgeon, the quality of the texts was lower (**Table 6**).

DISCUSSION

Today, it's very easy to search the internet for information, and patients often turn to the internet for research about their illnesses. However, there's no data on how accurate this habit is in accessing HPV-related information. This study aimed to determine the quality and readability of texts accessed by health readers on HPV-related websites.

It is known that there are over 100.000 websites providing health-related information.[18] With the increasing use of the Internet and the increase in health information on the Internet, it is seen that the impact on those who seek health information online on their medical decision making is also increasing.[19]

One of the most searched about topics on the internet in Turkiye is for health-related information at a rate of 65.9%.[3,4] While the internet offers tremendous opportunities for sharing information, it also spreads misinformation.[20]

To evaluate the quality of texts on websites, scoring is done according to a set of quality indicators. Today, there are many quality assessment tools.[6] JAMA and DISCERN are among the frequently used quality assessment tools today. In this study, websites about HPV were examined using these two tools. Although it was seen that websites prepared by obstetricians and gynecologists and obstetrician and gynecology associations were of "medium quality" according to both methods, the websites were of "poor quality" when considered in general.

An internet search can influence people's choices about whether or not to receive health care, leading to low-quality health information or content that can easily mislead or worry people.[21] Attention should be paid to health information of varying quality. There are many studies that have carried out quality evaluations on health-related websites using JAMA and DISCERN. Maheshwari et al.[22] examined 60 websites on postpartum pelvic health and reported that the general quality of health information available on the websites was of poor quality.

In another study using DISCERN and JAMA, Jain et al.[23] searched websites about 55 hysterectomy operations and found that the quality level of the online information about hysterectomy was sufficient, but the readability level of these texts was quite low. Although quality assessment tools can show whether a website contains reliable information, they do not predict its readability.[24]

Readability formulas and analysis tools are developed to determine the readability level or difficulty level of a text. In order for texts containing information about health to appeal

Table 6. Comparision of JAMA and DISCERN

		Mean	SD	Median	Minimum	Maximum	p-value
JAMA	Obstetrician and gynecologist	1.5	0.8	1.0	0.0	3.0	0.001
	Gynecological oncology surgeon	1.3	0.8	1.0	0.0	3.0	
DISCERN	Obstetrician and gynecologist	36.4	7.6	37.0	18.0	65.0	<0.001
	Gynecological oncology surgeon	33.0	6.9	33.0	17.0	63.0	

JAMA: Journal of the American Medical Association, SD: Standard deviation

to the general health readers, it should be aimed to reach a readability level between at least the fourth and sixth grades.[25] The use of readability formulas is increasing in order to reach the targeted readability levels in written health information materials, drug package inserts, and consent forms.[26,27]

In this study, in addition to evaluating the quality level of the websites, the readability of the websites was also examined, and the Ateşman and Bezirci-Yılmaz formulas were used to evaluate the readability of the Turkish texts. It was found that these texts were moderately readable according to the Ateşman readability formula, and according to the Bezirci-Yılmaz readability formula, they were readable with an education level of 12 years, that is, approximately at the undergraduate level.

In this case, it was seen that none of the websites included herein met their aims. In other studies that were conducted on the readability of websites, difficult texts were frequently encountered in terms of readability.[28] The readability level in the texts mainly prepared by the obstetrics and gynecology associations for patient education was not suitable for patients, and it is recommended to make appropriate arrangements to improve health literacy.[29] In the current study, it was seen that there were a few websites that provided information about HPV prepared by associations. It is noteworthy that this information, prepared by a small number of associations, received higher quality scores than the other groups.

When women first learn about their HPV carrier status, their susceptibility to anxiety and depression increases. With an appropriate treatment plan and psycho-social support, their anxiety decreases over time.[30] In this case, HPV-related websites with low-quality information and poor readability may increase the anxiety and depression of HPV-positive women and hinder the treatment and follow-up processes.

Studies on the readability and quality of information on health-related websites have been increasing in recent years, but no studies could be found investigating the quality and readability of health websites about HPV. Herein, it was observed that there was an inconsistency between the quality assessments of the two independent researchers. It was seen that the gynecological oncology surgeon scored lower than the obstetrician and gynecologist. It is believed that the reason for this was that surgeons, such as the gynecological oncology surgeon herein, have more up-to-date approaches.

Limitations

Health-related information from the internet can be obtained not only from written texts, but also from various media sources, such as YouTube and social media. In this study, the information obtained from such environments was not evaluated, and there were also factors other than the patient's education level that may affect readability, such as the font and font size. Such factors were not used in the evaluation herein. These conditions can be stated as the limitations of the study.

CONCLUSION

In this study, the health websites prepared on HPV were found to be of insufficient quality and difficult to read. It can be said that the readability and quality of Internet health information needs to be improved.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study has been approved by the Clinical Researches Ethics Committee of Süleyman Demirel University (Date: 07.07.2022, Decision No: 190).

Informed Consent

Since the work information was obtained from websites, informed consent is not required.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

Acknowledgement

We would like to thank Assoc, Prof. Erhan Okuyan for his contribution.

REFERENCES

1. Anderson JG, Rainey MR, Eysenbach G. The impact of CyberHealthcare on the physician-patient relationship. *J Med Syst*. 2003;27(1):67-84. doi:10.1023/a:1021061229743
2. Richards B, Colman AW, Hollingsworth RA. The current and future role of the Internet in patient education. *Int J Med Inform*. 1998;50(1-3):279-285. doi:10.1016/s1386-5056(98)00083-5
3. Bayrak A, Koçyiğit M, Aksu HS, Koçyiğit A. Z kuşağının sosyal medya bağımlılığı ve sosyotelist davranış ilişkisi: Konya örneği. *Akdeniz Üniversitesi İletişim Fakültesi Dergisi*. 2024;46:115-132. doi:10.31123/akil.1538072
4. Mansur F, Ciğerci K. Siberkondri ve E-sağlık okuryazarlığı arasındaki ilişki. *Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi*. 2022;11(1):11-21. doi:10.37989/gumussagbil.961457
5. Fox S. The engaged E-patient population. *Pew Internet Am Life Proj*. 2008;1-4.
6. Fahy E, Hardikar R, Fox A, Mackay S. Quality of patient health information on the internet: reviewing a complex and evolving landscape. *Australas Med J*. 2014;7(1):24-28. doi:10.4066/AMJ.2014.1900
7. Oxman AD, Guyatt GH, Cook DJ, Jaeschke R, Heddle N, Keller J. An index of scientific quality for health reports in the lay press. *J Clin Epidemiol*. 1993;46(9):987-1001. doi:10.1016/0895-4356(93)90166-x
8. Ahmad F, Hudak PL, Bercovitz K, Hollenberg E, Levinson W. Are physicians ready for patients with internet-based health information? *J Med Internet Res*. 2006;8(3):e22. doi:10.2196/jmir.8.3.e22
9. Manini I, Montomoli E. Epidemiology and prevention of human papillomavirus. *Ann Ig*. 2018;30(4 Supple 1):28-32. doi:10.7416/ai.2018.2231
10. Szymonowicz KA, Chen J. Biological and clinical aspects of HPV-related cancers. *Cancer Biol Med*. 2020;17(4):864-878. doi:10.20892/j.issn.2095-3941.2020.0370

11. Galeshi M, Shirafkan H, Yazdani S. et al. Reproducvtive health needs of women with human papillomavirus (HPV): a systematic review. *Research Square*. 2022. doi:10.21203/rs.3.rs-1410186/v1
12. Şahin F, Aydın E, Büyükkaya Öcal EU, Özdemir S, Kasapoğlu AM, Akbayır Ö. Evaluation of colposcopy and LEEP results performed in gynecology and gynecological oncology surgery services. *Eur J Gynaecol Oncol*. 2024;45(1):116. doi:10.22514/ejgo.2023.071
13. Atesman E. Measuring readability in Turkish. *AU Tömer Lang J*. 1997; 58(2):171-174.
14. Bezirci Y, Yılmaz A. A software library for measurement of readability of texts and a new readability metric for Turkish. *DEÜ FMD*. 2010;12(3):49-62.
15. Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewor--Let the reader and viewer beware. *JAMA*. 1997;277(15): 1244-1245. doi:10.1001/jama.1997.03540390074039
16. Khazaal Y, Chatton A, Cochand S, et al. Brief DISCERN, six questions for the evaluation of evidence-based content of health-related websites. *Patient Educ Couns*. 2009;77(1):33-37. doi:10.1016/j.pec.2009.02.016
17. Som R, Gunawardana NP. Internet chemotherapy information is of good quality: assessment with the DISCERN tool. *Br J Cancer*. 2012;107(2): 403. doi:10.1038/bjc.2012.223
18. Köhler C, Darmoni SJ, Mayer MA, Roth-Berghofer T, Fiene M, Eysenbach G. MedCIRCLE--the collaboration for internet rating, certification, labelling, and evaluation of health information. *Int J Heal Care Eng*. 2002;10(6):515. doi:10.3233/978-1-60750-939-4-667
19. Chen YY, Li CM, Liang JC, Tsai CC. Health information obtained from the internet and changes in medical decision making: questionnaire development and cross-sectional survey. *J Med Internet Res*. 2018;20(2): e47. doi:10.2196/jmir.9370
20. Wang Y, McKee M, Torbica A, Stuckler D. systematic literature review on the spread of health-related misinformation on social media. *Soc Sci Med*. 2019;240(September):112552. doi:10.1016/j.socscimed.2019.112552
21. Levine DM, Mehrotra A. Assessment of diagnosis and triage in validated case vignettes among nonphysicians before and after internet search. *JAMA Netw Open*. 2021;4(3):1-12. doi:10.1001/jamanetworkopen.2021.3287
22. Maheshwari D, Kunycky C, Jia X, Tangada A, Leung K, Solomon ER. Patient-focused websites related to postpartum pelvic floor health: a DISCERN quality analysis. *Female Pelvic Med Reconstr Surg*. 2022;28(4): 240-243. doi:10.1097/SPV.0000000000001101
23. Jain M, Chkipov P, Stacey D, Posner G, Bacal V, Chen I. Online patient information for hysterectomies: an environmental scan of readability and quality. *J Obstet Gynaecol Canada*. 2021;43(5):667. doi:10.1016/j.jogc.2021.02.061
24. Sobota A, Ozakinci G. The quality and readability of online consumer information about gynecologic cancer. *Int J Gynecol Cancer*. 2015;25(3): 537-541. doi:10.1097/IGC.0000000000000362
25. Matthews T, Sewell J. Executive summary: excerpt from CSG's state official's guide to health literacy. The Council of State Governments. United States of America. 2002.
26. Wang LW, Miller MJ, Schmitt MR, Wen FK. Assessing readability formula differences with written health information materials: application, results, and recommendations. *Res Soc Adm Pharm*. 2013; 9(5):503-516. doi:10.1016/j.sapharm.2012.05.009
27. Ay IE, Doğan M. An evaluation of the comprehensibility levels of ophthalmology surgical consent forms. *Cureus*. 2021;13(7):10-15. doi:10.7759/cureus.16639
28. Walsh TM, Volsko TA. Readability assessment of internet-based consumer health information. *Respir Care*. 2008;53(10):1310-1315.
29. Fahimuddin FZ, Sidhu S, Agrawal A. Reading level of online patient education materials from major obstetrics and gynecology societies. *Obstet Gynecol*. 2019;133(5):987-993. doi:10.1097/AOG.0000000000003214
30. Uzun SB, Sakin Ö, Hüseyin Ç, Şimsek EE. The effects of HPV test on anxiety, emotion and depression in women. *J Acad Res Med*. 2020;10(2): 149-154. doi:10.4274/jarem.galenos.2020.3106

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1738996

Anatolian Curr Med J. 2025;7(5):569-575

Spread through air spaces as a prognostic factor in resected non-small cell lung cancer

Gül Sema Yıldıran Keskin[1], Gizem Yıldırım[1], Volkan Aslan[1], Elçin Kadan[2], Ersin Sapmaz[3], Musa Barış Aykan[1], İsmail Ertürk[1], Nuri Karadurmuş[1]

[1]*Department of Medical Oncology, Gülhane Training and Research Hospital, University of Health Sciences, Ankara, Turkiye*
[2]*Department of Medical Pathology, Gülhane Training and Research Hospital, University of Health Sciences, Ankara, Turkiye*
[3]*Department of Thoracic Surgery, Gülhane Training and Research Hospital, University of Health Sciences, Ankara, Turkiye*

Cite this article as: Yıldıran Keskin GS, Yıldırım G, Aslan V, et al. Spread through air spaces as a prognostic factor in resected non-small cell lung cancer. *Anatolian Curr Med J.* 2025;7(5):569-575.

Received: 10.07.2025 ◆ **Accepted**: 05.08.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: To evaluate the prognostic impact of spread through air spaces (STAS) in patients with resected non-small cell lung cancer (NSCLC) and to investigate its association with clinicopathological features and survival outcomes.

Methods: A retrospective analysis was conducted on 207 patients with pathological stage IA–IIIA NSCLC who underwent curative-intent surgery between 2018 and 2024. STAS was defined as the presence of micropapillary clusters, solid nests, or single tumor cells within alveolar spaces beyond the main tumor. Patients were categorized as STAS-positive or STAS-negative. Disease-free survival (DFS) and overall survival (OS) were estimated using the Kaplan–Meier method. Univariate and multivariate Cox regression models were used to assess prognostic factors.

Results: STAS was identified in 57% of patients. STAS positivity was significantly associated with poor differentiation, a higher rate of lymphovascular invasion, and increased use of adjuvant chemotherapy. Median DFS was 29.9 months in STAS-positive patients but was not reached in STAS-negative patients ($p<0.001$). In multivariable analysis, STAS independently predicted shorter DFS (HR: 2.38; 95% CI: 1.34–4.23; $p=0.003$). No statistically significant association was found between STAS and OS ($p=0.079$).

Conclusion: STAS is an independent adverse prognostic factor for DFS in resected NSCLC. Its presence should be considered in prognostic evaluation and surgical planning, particularly in patients with early-stage disease.

Keywords: Non-small cell lung cancer, STAS, disease-free survival, prognostic factors, surgery

INTRODUCTION

In recent years, the implementation of low-dose computed tomography screening programs in some countries has led to a declining trend in lung cancer incidence and mortality. However, lung cancer remains the most frequently diagnosed malignancy and the leading cause of cancer-related death worldwide. According to GLOBOCAN 2022 estimates, there were 2.48 million new cases and 1.82 million deaths globally, reflecting the ongoing burden of this disease.[1] Despite improvements in early detection, non-small cell lung cancer (NSCLC) continues to exhibit aggressive biological behavior even in early-stage and operable disease. SEER data indicate that the 5-year recurrence rate after surgical resection for NSCLC ranges from approximately 30% to nearly 70%.[2] These findings underscore the need for robust prognostic stratification at the time of diagnosis, particularly in early-stage disease.

Several prognostic factors associated with unfavorable outcomes after surgical resection have been identified, including pleural invasion, lymphovascular invasion (LVI), poor tumor differentiation, wedge resection, and unknown lymph node status.[3] In this context, spread through air spaces (STAS) which is described by the World Health Organization (WHO) in 2015, initially in lung adenocarcinoma. Since then, STAS has garnered increasing attention for its potential prognostic significance in NSCLC.[4,5] STAS is defined as the presence of tumor cells within the adjacent alveolar parenchyma beyond the edge of the main tumor, detectable microscopically in lung cancer specimens. It has been reported in approximately 15% to 73% of surgically resected lung cancers and is associated with poor prognosis.[6-10] This adverse prognostic association has been consistently observed across all major histological subtypes of lung cancer studied, including adenocarcinoma, squamous cell carcinoma, small cell carcinoma and others.[11,12]

While many of the traditional prognostic factors reflect tumor burden or invasiveness, they may not fully capture

Corresponding Author: Gül Sema Yıldıran Keskin, gulsemayildiran@gmail.com

microscopic patterns of tumor spread that directly influence the risk of recurrence. STAS represents a distinct pattern of tumor dissemination through alveolar spaces, independent of vascular or lymphatic spread.[13] Importantly, its presence has been associated with worse outcomes even among patients with otherwise favorable pathological features. These observations suggest that STAS may serve as an independent prognostic marker and a valuable criterion for informing surgical decisions, particularly when considering sublobar resections.[14,15] Accordingly, increasing efforts have been made to incorporate STAS into prognostic algorithms and clinical decision-making in resectable NSCLC.

Therefore, the aim of this study was to evaluate the prognostic significance of STAS in patients with resected NSCLC, and to investigate its association with clinicopathological factors and survival outcomes in a real-world, single-center cohort.

METHODS

Ethics

The study has been approved by the Scientific Researches Ethics Committee of Gülhane Training and Research Hospital (Date: 06.05.2025, Decision No: 2025-275). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Study Design and Participants

This retrospective single-center study included 207 patients who underwent curative-intent surgical resection for pathological stage IA–IIIA NSCLC between 2018 and 2024 at Gülhane Training and Research Hospital. Eligible patients were selected based on the availability of pathological STAS assessment and complete clinical and follow-up data.

Demographic variables (age, sex), smoking history, Eastern Cooperative Oncology Group performance status (ECOG PS), tumor characteristics (histological subtype, differentiation, tumor size), pathological features (LVI, perineural invasion [PNI], visceral pleural invasion [VPI]), and molecular markers (PD-L1 expression, mutational status) were recorded. Radiological staging with positron emission tomography/computed tomography (PET/CT), type of surgical intervention (wedge resection, lobectomy), and lymph node evaluation status were also collected. Surgical approach and extent of resection were determined based on tumor size, location, and patient comorbidities. Of the total 207 patients, 9 underwent wedge resection and 2 underwent other forms of sublobar resection due to high comorbidity burden and/or small tumor size. The remaining patients were treated with lobectomy or pneumonectomy accompanied by systematic mediastinal and hilar lymph node sampling.

Prognostic outcomes included overall survival (OS) and disease-free survival (DFS). OS was defined as the time from the date of surgery to death from any cause, and DFS as the time from surgery to recurrence or death. The primary objective was to evaluate the prognostic impact of STAS on survival outcomes.

Histopathological Assessment

Histopathological evaluation was performed on resection specimens obtained during definitive surgery. STAS assessment was performed exclusively on resection specimens; no frozen sections or preoperative biopsies were used. All specimens were fixed in 10% neutral-buffered formalin and embedded in paraffin, and processed according to routine histological procedures. Four-micron thick sections were stained with hematoxylin and eosin (H&E) for microscopic examination.

STAS was defined in accordance with the 2015 WHO classification of lung tumors as the presence of tumor cells either as micropapillary clusters, solid nests, or single tumor cells within alveolar spaces beyond the edge of the main tumor. To minimize misinterpretation due to histological artifacts, particular attention was paid to exclude free-floating cell strips or fragmented clusters lacking alveolar attachment, as these are often considered artifacts from specimen handling rather than true STAS. All histopathological assessments were performed by an experienced thoracic pathologist as part of routine diagnostic practice at our institution. A second, independent pathology review was not conducted for this retrospective study. Tumor histological subtype, grade, and presence of lymphovascular or pleural invasion were also assessed and documented.

Statistical Analysis

Clinicopathological variables were compared between STAS-positive and STAS-negative groups using Pearson's Chi-squared test or Fisher's exact test, as appropriate. Descriptive statistics were presented as numbers and percentages for categorical variables.

OS and DFS were estimated using the Kaplan–Meier method, and survival differences between STAS subgroups were assessed using the log-rank test. The median follow-up time was calculated using the reverse Kaplan–Meier method. Univariate Cox proportional hazards regression was used to identify factors associated with OS and DFS. Variables with a p-value <0.05 in the univariate analysis were included in the multivariate Cox regression model to identify independent prognostic factors. Hazard ratios (HRs) and 95% confidence intervals (CIs) were calculated.

All statistical analyses were conducted using IBM SPSS Statistics for Windows, Version 25 (IBM Corp., Armonk, NY, USA). A p-value <0.05 was considered statistically significant.

RESULTS

Clinicopathological Characteristics of the Patients

The clinicopathological characteristics of the 207 patients are summarized in **Table 1**. The median age was 64 years (min-max: 39–82), and 78.7% were younger than 70 years. Most patients were male (78.3%, n=162). Regarding smoking history, 82.2% were active or former smokers, and 17.8% had never smoked. STAS positivity was identified in 118 patients (57%), whereas 89 patients (43%) were STAS-negative.

Table 1. Demographic and clinicopathological parameters according to STAS status

Variables	All patients n=207	STAS (+) n=118	STAS (-) n=89	p value
Age, median (min-max)	64 (39-82)	64 (41-82)	62 (39-81)	0.37
<70 years, n (%)	163 (78.7%)	91 (77.1%)	72 (80.9%)	0.61
≥70 years, n (%)	44 (21.3%)	27 (22.9%)	17 (19.1%)	
Sex, n (%)				
Female	45 (21.7%)	20 (16.9%)	25 (28.1%)	0.06
Male	162 (78.3%)	98 (83.1%)	64 (71.9%)	
Smoking status, n (%)	35 (17.8%)	19 (17%)	16 (18.8%)	
Never smoker	162 (82.2%)	93 (83%)	69 (81.2%)	0.86
Active/former-smoker				
Type of surgery, n (%)				
Wedge	11 (5.3%)	3 (2.5%)	8 (9%)	
Lobectomy	185 (89.4%)	111 (94.1%)	74 (83.1%)	0.038
Pneumonectomy	11 (5.3%)	4 (3.4%)	7 (7.9)	
TNM stage, n (%)				
1A	74 (35.7%)	31 (26.3%)	43 (48.3%)	
1B	48 (23.2%)	24 (20.3%)	24 (27%)	
2A	13 (6.3%)	10 (8.5%)	3 (3.4%)	0.001
2B	34 (16.4%)	2.5 (2.2%)	9 (10.1%)	
3A	38 (18.4%)	28 (23.7%)	10 (11.2%)	
Pathological type, n (%)				
Adenocarcinoma	126 (60.9%)	67 (56.8%)	59 (66.3%)	
SqCC	70 (33.8%)	46 (39%)	24 (27%)	0.17
LC-NEC	11 (5.3%)	5 (4.2%)	6 (6.7%)	
Histological grading, n (%)				
Good	21 (10.9%)	7 (6.4%)	14 (17.1%)	
Moderate	124 (64.6%)	69 (62.7%)	55 (67.1%)	0.009
Poor	47 (24.5%)	34 (30.9%)	13 (15.9%)	
LVI, n (%)				
Presence	85 (41.1%)	56 (47.5%)	29 (32.6%)	0.033
Absence	122 (58.9%)	62 (52.5%)	60 (67.4%)	
VPI, n (%)				
Presence	61 (29.5%)	38 (32.2)	23 (25.8%)	0.358
Absence	146 (70.5%)	80 (67.8%)	66 (74.2%)	
Adenocarcinoma subtypes, n (%)	126 (100%)			
Lepidic	68 (54%)	30 (44.8%)	38 (64.4%)	0.032
Acinar	90 (71.4%)	46 (68.7%)	44 (74.6%)	0.554
Solid	54 (42.9%)	33 (49.3%)	21 (35.6%)	0.122
Papillary	47 (39.3%)	24 (35.8%)	23 (39%)	0.854
Micropapillary	19 (15.1%)	13 (19.4%)	6 (10.2%)	0.212
PDL1, n (%)	78 (100%)			
<1	53 (67.9%)	34 (67.3%)	16 (69.6%)	0.843
≥1	25 (32.1%)	18 (32.7%)	7 (30.4%)	
Mutational status, n (%)	76 (100%)			
EGFR	8 (10.5%)	5 (9.6%)	3 (12.5%)	
KRAS	5 (6.6%)	4 (7.7%)	1 (4.2%)	0.793
None	63 (82.9%)	43 (82.7%)	20 (83.3%)	
Adjuvant CT, n (%)	141(68.1%)	88 (74.6%)	53 (59.6%)	
Presence	66 (31.9%)	30 (25.4%)	36 (40.4%)	0.024
Absence				
Progression	81 (39.1%)	62 (52.5%)	19 (21.3%)	<0.001
Exitus	49 (23.7%)	33 (28%)	16 (18%)	0.102

STAS: Spread through air spaces, LVI: Lymphovascular invasion, VPI: Visceral pleural invasion, DFS: Disease-free survival, OS: Overall survival, CT: Chemotherapy, EGFR: Epidermal growth factor receptor; KRAS: Kirsten rat sarcoma virus, PD-L1: Programmed death ligand 1, SqCC: Squamous cell carcinoma, LC-NEC: Large cell neuroendocrine carcinoma, TNM: Tumor-node-metastasis (staging system).
Note: Percentages in Table 1 are calculated column-wise. Chi-square test was used for comparisons between STAS-positive and STAS-negative groups.

Lobectomy was the most common surgical approach (89.4%, n=185). Pathological staging revealed stage IA in 35.7% of patients, IB in 23.2%, IIA in 6.3%, IIB in 16.4%, and IIIA in 18.4%. Stage IA was significantly more common among STAS-negative patients (48.3% vs. 26.3%, p=0.001).

Histologically, 60.9% of tumors were adenocarcinomas (n=126), 33.8% were squamous cell carcinomas (n=70), and 5.3% were other subtypes, including large cell neuroendocrine carcinoma and atypical carcinoid tumors (p=0.17). In terms of differentiation, 10.9% were well-differentiated, 64.6% were moderately differentiated, and 24.5% were poorly differentiated. Poor differentiation was significantly more common in the STAS-positive group (30.9% vs. 15.9%, p=0.009). LVI was observed in 41.1% (n=85) of cases, with a higher prevalence in STAS-positive patients (47.5% vs. 32.6%, p=0.033). VPI was observed in 29.5% of patients, with no significant difference between STAS groups (p=0.358).

Molecular status (EGFR, ALK, ROS1, KRAS) was available for 76 patients (36.7%), and PD-L1 expression was assessed in 78 patients (37.7%). Among those tested, EGFR mutations were detected in 8 patients (10.5%) and KRAS mutations in 5 patients (6.6%). All patients tested for ALK and ROS1 rearrangements were negative. PD-L1 expression of ≥1% was observed in 25 patients (32.1%). The distribution of these

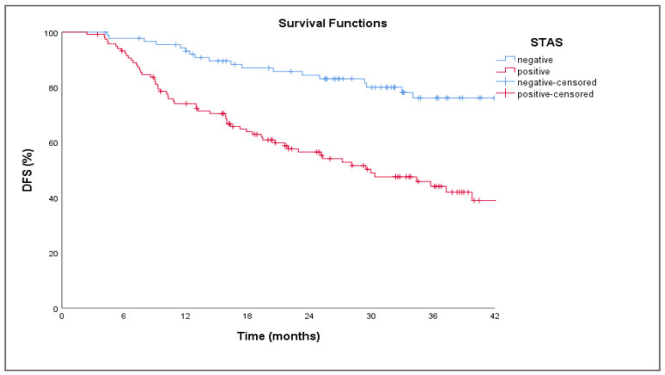
molecular alterations did not significantly differ between STAS-positive and STAS-negative groups (p>0.05 for all comparisons).

Adjuvant chemotherapy was administered to 68.1% of patients and was significantly more common in the STAS-positive group (74.6% vs. 59.6%, p=0.024). No patients received induction therapy before surgery.

Impact of STAS on Survival Outcomes

The median follow-up duration was 36.5 months. During this period, the estimated median DFS was 51.7 months, and the median OS was 90.6 months. However, CIs could not be calculated because of the high proportion of censored cases.

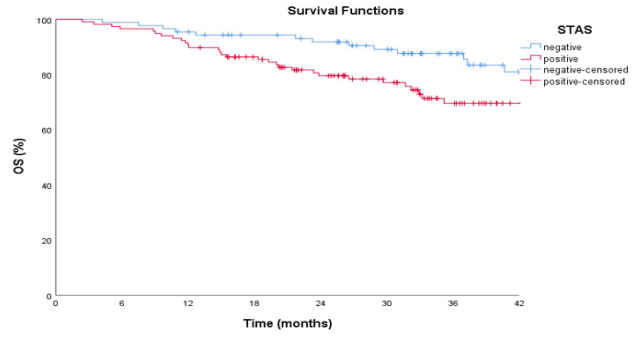
STAS-positive patients had significantly worse survival outcomes. The median DFS in the STAS-positive group was 29.9 months (95% CI: 20.6–39.4), while it was not reached in the STAS-negative group (p<0.001). Similarly, the median OS was 62.3 months (95% CI: 35.2–89.4) in STAS-positive patients and was not reached in the STAS-negative group (p=0.017). Kaplan-Meier survival curves for DFS and OS according to STAS status are presented in **Figure 1, 2**.



No. at risk

Months	0	6	12	18	24	30	36	42
STAS-negative	88	85	79	73	67	62	57	51
STAS-positive	117	114	107	101	95	89	83	78

Figure 1. Kaplan-Meier curves for disease-free survival (DFS) according to STAS status. Median DFS was 29.9 months in STAS-positive patients and not reached in STAS-negative patients (log-rank p<0.001).
STAS: Spread through air spaces



No. at risk

Months	0	6	12	18	24	30	36	42
STAS-negative	88	87	83	80	76	70	63	57
STAS-positive	117	115	112	108	104	98	93	87

Figure 2. Kaplan-Meier curves for overall survival (OS) according to STAS status. STAS-positive patients had significantly shorter median OS compared to STAS-negative patients (62.3 vs. not reached; p=0.017).
STAS: Spread through air spaces

In univariable analysis, in addition to STAS positivity, several clinicopathological variables were significantly associated with shorter DFS. Poor tumor differentiation (HR: 2.36; 95% CI: 1.56-3.58; p<0.001), presence of LVI (HR: 1.84; 95% CI: 1.19-2.86; p=0.006) and VPI (HR: 1.90; 95% CI: 1.21-2.97; p=0.005) were also significantly associated with worse DFS. Patients who received adjuvant chemotherapy had shorter DFS compared to those who did not (HR: 1.87; 95% CI: 1.12-3.14; p=0.016). In contrast, age, sex, smoking status, type of surgery, and histologic subtype were not significantly associated with DFS. Three-year DFS rates differed significantly according to TNM stage, with stage IA showing the most favorable rate (73.6%). **Table 2** summarizes the univariable analyses for DFS and OS.

In the multivariable Cox regression model, STAS positivity remained an independent predictor of disease recurrence (HR: 2.38; 95% CI: 1.34–4.23; p=0.003). VPI was also significantly associated with shorter DFS (HR: 1.79; 95% CI: 1.03–3.10; p=0.038), as was poor differentiation (HR: 3.53; 95% CI: 1.17–10.62; p=0.025). TNM stage, LVI, and adjuvant chemotherapy were not statistically significant in this model.

In the multivariable model for OS, age ≥70 years was independently associated with poorer survival (HR: 2.50; 95% CI: 1.27–4.92; p=0.008). Although STAS positivity (HR: 1.86; 95% CI: 0.93–3.72; p=0.079) and poor differentiation (HR: 3.05; 95% CI: 0.84–11.03; p=0.090) demonstrated a trend toward poorer OS, these did not reach statistical significance. Other variables, including TNM stage, LVI, and adjuvant chemotherapy, were not independently associated with OS. Multivariable Cox regression analyses for DFS and OS are presented in **Table 3**.

DISCUSSION

In this retrospective cohort study, we demonstrated that the presence of STAS was significantly associated with several adverse clinicopathological features, including higher pathological stage, poor tumor differentiation, and LVI. In multivariable analyses, STAS emerged as an independent predictor of DFS (HR: 2.38; 95% CI: 1.34–4.23; p=0.003). These findings suggest that STAS is not merely a histopathological observation but may reflect a more aggressive tumor biology, contributing to an increased risk of recurrence even in early-stage disease.

The incidence of STAS in our cohort was 57%, which is within the higher range reported in the literature. Previous studies, such as those by Toyokawa et al.[10] and Gutierrez-Sainz et al.,[16] reported STAS positivity rates of 71.2% and 73%, respectively. These elevated rates were largely attributed to the inclusion of patients with more advanced disease stages (stage II and III) in their cohorts. Similarly, our study also revealed a significant association between STAS positivity and higher pathological stage; stage IA disease was notably more frequent among STAS-negative patients (48.3% vs. 26.3%, p=0.001). These results support earlier findings that STAS is closely linked to tumor aggressiveness and may reflect more advanced tumor biology.

Table 2. Univariable analysis for DFS and OS

		Univariable for DFS		Univariable for OS	
Variables	**3 year DFS (%)**	**HR (95% CI)**	**p value**	**HR**	**p value**
<70 years	59.2	Reference	0.509	Reference	0.003
>70 years	54.4	1.19(0.70-2.04)		2.48 (1.37-4.50)	
Female	62.6	Reference	0.296	Reference	0.358
Male	56.9	1.34 (0.76-2.36)		1.42 (0.66-3.05)	
Never smoker	51.2	Reference	0.800	Reference	0.720
Active/former smoker	58.3	1.08 (0.59-1.96)		1.15 (0.51-2.58)	
Wedge resection	60.0	Reference	0.815	Reference	0.140
Lobectomy	58.8	0.85 (0.31-2.35)	0.766	0.61 (0.28-1.71)	0.347
Pneumonectomy	48.5	1.12 (0.30-4.19)	0.863	1.45 (0.39-5.45)	0.575
TNM stage					
1A	73.6	Reference	<0.001	Reference	0.003
1B	69.7	0.97 (0.48-1.96)	0.945	0.52 (0.18-1.48)	0.226
2A	41.0	2.73 (1.15-6.48)	0.022	3.78 (1.42-10.04)	0.008
2B	34.2	2.76 (1.47-5.18)	0.002	2.00 (0.89-4.46)	0.091
3A	40.9	3.00 (1.60-5.51)	<0.001	2.45 (1.13-5.31)	0.023
LVI					
Absent	68	Reference	0.006	Reference	0.030
Present	44.6	1.84 (1.19-2.86)		1.87 (1.06-3.31)	
VPI					
Absent	65.3	Reference	0.005	Reference	0.144
Present	42.3	1.90 (1.21-2.97)		1.54 (0.86-2.78)	
STAS					
Absent	76.1	Reference	<0.001	Reference	0.019
Present	44.1	3.60 (2.13-6.06)		2.10 (1.13-3.90)	
Differantiation					
Good	77.0	Reference	<0.001	Reference	0.030
Intermediate	63.0	2.13 (0.76-5.98)	0.147	1.36 (0.41-4.53)	0.613
Poor	38.9	5.19(1.80-14.95)	0.002	3.53 (0.88-10.66)	0.077
Histological subtype					
Adenocancer	58.5	Reference	0.693	Reference	0.350
Squamous cell cancer	58.9	0.86 (0.53-6.40)	0.562	1.66 (0.90-3.03)	0.099
Others	50.9	1.27 (0.50-3.21)	0.601	3.30 (1.24-8.77)	0.016
Adjuvant CT					
Absent	72.7	Reference	0.016	Reference	0.020
Present	51.5	1.87 (1.12-3.14)		2.36 (1.14-4.89)	

STAS: Spread through air spaces, LVI: Lymphovascular invasion, VPI: Visceral pleural invasion, DFS: Disease-free survival, OS: Overall survival, HR: Hazard ratio, CI: Confidence interval, CT: Chemotherapy, TNM: Tumor-node-metastasis (staging system)

Table 3. Multivariable Cox regression analysis for DFS and OS

	DFS		OS	
Variables	**HR (95% CI)**	**p value**	**HR (95% CI)**	**p value**
Age <70 vs >70			2.50 (1.27-4.92)	0.008
TNM stage				
1A vs	Reference	0.337	Reference	0.106
1B	0.54 (0.22-1.36)	0.198	0.25 (0.07-0.89)	0.032
2A	0.99 (0.35-2.79)	0.994	1.29 (0.40-4.09)	0.666
2B	1.08 (0.41-2.88)	0.867	0.59 (0.19-1.81)	0.364
3A	1.26 (0.49-3.20)	0.625	0.98 (0.34-2.77)	0.976
LVI (absent vs present)	1.49 (0.86-2.56)	0.148	1.79 (0.89-3.58)	0.100
VPI (absent vs present)	1.78 (1.03-3.10)	0.038		
STAS (absent vs present)	2.38 (1.34-4.23)	0.003	1.85 (0.93-3.71)	0.079
Differantiation	Reference	0.005	Reference	0.036
Good vs intermediate	1.58 (0.55-4.50)	0.389	1.30 (0.38-4.41)	0.667
Good vs poor	3.53 (1.17-10.6)	0.025	3.04 (0.84-11.03)	0.090
Adjuvant CT	1.08 (0.49-2.35)	0.844	2.20 (0.81-5.92)	0.119

STAS: Spread through air spaces, LVI: Lymphovascular invasion, VPI: Visceral pleural invasion, DFS: Disease-free survival, OS: Overall survival, HR: Hazard ratio, CI: Confidence interval, CT: Chemotherapy, TNM: Tumor-node-metastasis (staging system).

In addition to pathological stage, STAS was also associated with tumor histologic subtype and differentiation.[17,18] In our study, STAS was more frequently observed in poorly differentiated tumors, further supporting its link to aggressive tumor behavior. Among adenocarcinoma subtypes, the absence of a lepidic component was associated with the presence of STAS. Consistent with our findings, studies by Xie et al.[19] and Cao et al.[20] have also demonstrated that the solid growth pattern is significantly correlated with STAS positivity.

The prognostic importance of STAS is increasingly recognized in the literature. In a recent study by Chen et

al.[21] focusing on stage IA lung adenocarcinoma ≤2 cm, STAS was found in 43.4% of cases, and its association with adverse pathological features, including poor differentiation and LVI, was confirmed. Importantly, STAS remained an independent prognostic factor for OS in multivariate analysis. Similarly, another study showed that STAS was associated with worse 5-year DFS and OS in stage IB (T2aN0) NSCLC patients, emphasizing its impact even in early stage disease.[6] A large-scale analysis by the International Association for the Study of Lung Cancer (IASLC) included 4.061 resected NSCLC cases and found STAS in 22.9% of tumors. In this study, STAS was independently associated with both DFS and OS, regardless of disease stage. Based on these data, the authors suggested that STAS should be included in the 9th TNM classification, along with factors such as visceral pleural and LVI.[5] According to our findings, STAS is an independent prognostic marker for DFS. While a trend toward worse OS was observed in STAS-positive patients, this association did not reach statistical significance in our multivariate analysis (HR: 1.86; 95% CI: 0.93–3.72; p=0.079). This may be explained by the relatively low number of events and the limited follow-up duration in our cohort, both of which could have reduced the statistical power to detect a significant difference. With longer follow-up and larger patient numbers, the prognostic impact of STAS on OS may become more evident.

Our study did not find a significant association between STAS and molecular markers such as PD-L1 expression or driver mutations. However, this should be interpreted cautiously, as molecular testing was unavailable in approximately two-thirds of patients. This reflects the national healthcare context in Turkiye during the study period, when molecular testing was not routinely performed for early-stage NSCLC due to lack of reimbursement and limited access to targeted therapies and immunotherapies. Consequently, molecular profiling was often limited to patients with recurrence, thereby reducing the ability to explore correlations between STAS and specific molecular alterations in the overall cohort.

The relationship between STAS and molecular alterations remains controversial in the literature. Lee et al.[17] and Gutierrez-Sainz et al.[16] found that EGFR mutations were less common in STAS-positive tumors, suggesting a potential inverse correlation. In contrast, Tian et al.[22] reported that STAS was more frequently observed in EGFR-mutant tumors and, demonstrated a significant association between ALK alterations and STAS positivity. Further insights were gained from a recent large-scale genomic profiling study by Ye et al.[23] who analysed 442 resected lung adenocarcinomas using next-generation sequencing. They found that EGFR mutations were significantly less frequent in STAS-positive tumors (52.5% vs. 69.7%, p<0.001), while TP53 mutations and ALK rearrangements were enriched in the STAS-positive group. Taken together, these conflicting and evolving data underline the complexity of STAS pathogenesis and highlight the need for further large-scale studies with standardized molecular profiling to clarify these associations.

In our study, lobectomy was more commonly performed in STAS-positive patients. The number of patients who underwent sublobar resection in our cohort was limited, which restricts our ability to draw definitive conclusions regarding the impact of STAS positivity on surgical outcomes in this subgroup. Several studies have reported that patients with STAS-positive tumors who undergo sublobar resections such as wedge resection or segmentectomy have higher recurrence rates and worse disease-free and OS compared to those who undergo lobectomy.[24,25] However, this association remains controversial. Kagimoto et al.[26] found that segmentectomy provided comparable oncologic outcomes to lobectomy in patients with stage IA lung adenocarcinoma with STAS, without increasing the risk of locoregional recurrence. Furthermore, some investigators have suggested that adjuvant therapy may be warranted in STAS-positive patients undergoing sublobar resections, particularly in early-stage disease, to mitigate the risk of recurrence.[27] Overall, these results highlight the importance of considering STAS as both a prognostic biomarker and a factor that may influence surgical decision-making; however, these conclusions should be interpreted with caution and require validation in larger studies.

Limitations

This study has several limitations. First, it was a retrospective, single-center analysis, which may have introduced selection bias and limited the generalizability of the results. Second, although the cohort size was moderate, the number of events and duration of follow-up may have been insufficient to detect statistically significant differences in OS. Third, molecular testing was not performed in a substantial proportion of patients, and as a result, our ability to comprehensively evaluate the association between STAS and specific molecular alterations was limited.

CONCLUSION

As a result, our findings confirm that the presence of STAS is significantly associated with adverse pathological features and independently predicts DFS in resected NSCLC. These results highlight the potential clinical utility of STAS as a prognostic biomarker particularly in early-stage disease. Future prospective studies with longer follow-up and comprehensive molecular profiling are needed to validate and expand these findings.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study has been approved by the Scientific Researches Ethics Committee of Gülhane Training and Research Hospital (Date: 06.05.2025, Decision No: 2025-275).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Filho AM, Laversanne M, Ferlay J, et al. The GLOBOCAN 2022 cancer estimates: data sources, methods, and a snapshot of the cancer burden worldwide. *Int J Cancer.* 2025;156(7):1336-1346. doi:10.1002/ijc.35278
2. Institute NC. Cancer stat facts: lung and bronchus cancer. Surveillance, Epidemiology, and End Results (SEER) Program, NCI. 28.06.2025, Accessed 28 June 2025, https://seer.cancer.gov/statfacts/html/lungb.html
3. NCCN. NCCN Clinical Practice Guidelines in Oncology: non-small cell lung cancer (Version 5.2025). National Comprehensive Cancer Network. 28 June 2025, Accessed 28 June 2025, https://www.nccn.org/professionals/physician_gls/pdf/nscl.pdf
4. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization Classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol.* 2015;10(9):1243-1260. doi:10.1097/jto.0000000000000630
5. Travis WD, Eisele M, Nishimura KK, et al. The International Association for the Study of Lung Cancer (IASLC) staging project for lung cancer: recommendation to introduce spread through air spaces as a histologic descriptor in the ninth edition of the TNM classification of lung cancer. Analysis of 4061 pathologic stage I NSCLC. *J Thorac Oncol.* 2024;19(7):1028-1051. doi:10.1016/j.jtho.2024.03.015
6. Chen D, Mao Y, Wen J, et al. Tumor spread through air spaces in non-small cell lung cancer: a systematic review and meta-analysis. *Ann Thorac Surg.* 2019;108(3):945-954. doi:10.1016/j.athoracsur.2019.02.045
7. Terada Y, Takahashi T, Morita S, et al. Spread through air spaces is an independent predictor of recurrence in stage III (N2) lung adenocarcinoma. *Interact Cardiovasc Thorac Surg.* 2019;29(3):442-448. doi:10.1093/icvts/ivz116
8. Ren Y, Xie H, Dai C, et al. Prognostic impact of tumor spread through air spaces in sublobar resection for 1A lung adenocarcinoma patients. *Ann Surg Oncol.* 2019;26(6):1901-1908. doi:10.1245/s10434-019-07296-w
9. Toki MI, Harrington K, Syrigos KN. The role of spread through air spaces (STAS) in lung adenocarcinoma prognosis and therapeutic decision making. *Lung Cancer.* 2020;146:127-133. doi:10.1016/j.lungcan.2020.04.026
10. Toyokawa G, Yamada Y, Tagawa T, et al. Significance of spread through air spaces in resected lung adenocarcinomas with lymph node metastasis. *Clin Lung Cancer.* 2018;19(5):395-400 e1. doi:10.1016/j.cllc.2018.04.002
11. Lee MA, Kang J, Lee HY, et al. Spread through air spaces (STAS) in invasive mucinous adenocarcinoma of the lung: incidence, prognostic impact, and prediction based on clinicoradiologic factors. *Thorac Cancer.* 2020;11(11):3145-3154. doi:10.1111/1759-7714.13632
12. Aly RG, Rekhtman N, Li X, et al. Spread through air spaces (STAS) is prognostic in atypical carcinoid, large cell neuroendocrine carcinoma, and small cell carcinoma of the lung. *J Thorac Oncol.* 2019;14(9):1583-1593. doi:10.1016/j.jtho.2019.05.009
13. Alvarez Moreno JC, Aljamal AA, Bahmad HF, et al. Correlation between spread through air spaces (STAS) and other clinicopathological parameters in lung cancer. *Pathol Res Pract.* 2021;220:153376. doi:10.1016/j.prp.2021.153376
14. Liu C, Wang YF, Wang P, et al. Predictive value of multiple imaging predictive models for spread through air spaces of lung adenocarcinoma: a systematic review and network meta-analysis. *Oncol Lett.* 2024;27(3):122. doi:10.3892/ol.2024.14255
15. Eguchi T, Kameda K, Lu S, et al. Lobectomy is associated with better outcomes than sublobar resection in spread through air spaces (STAS)-positive T1 lung adenocarcinoma: a propensity score-matched analysis. *J Thorac Oncol.* 2019;14(1):87-98. doi:10.1016/j.jtho.2018.09.005
16. Gutierrez-Sainz L, López-Muñoz S, Cruz-Castellanos P, et al. Retrospective analysis of the prognostic implications of tumor spread through air spaces in lung adenocarcinoma patients treated with surgery. *ESMO Open.* 2022;7(5):100568. doi:10.1016/j.esmoop.2022.100568
17. Lee JS, Kim EK, Kim M, Shim HS. Genetic and clinicopathologic characteristics of lung adenocarcinoma with tumor spread through air spaces. *Lung Cancer.* 2018;123:121-126. doi:10.1016/j.lungcan.2018.07.020
18. Hu SY, Hsieh MS, Hsu HH, et al. Correlation of tumor spread through air spaces and clinicopathological characteristics in surgically resected lung adenocarcinomas. *Lung Cancer.* 2018;126:189-193. doi:10.1016/j.lungcan.2018.11.003
19. Xie H, Su H, Zhu E, et al. Morphological subtypes of tumor spread through air spaces in non-small cell lung cancer: prognostic heterogeneity and its underlying mechanism. *Front Oncol.* 2021;11608353. doi:10.3389/fonc.2021.608353
20. Cao L, Jia M, Sun PL, Gao H. Histopathologic features from preoperative biopsies to predict spread through air spaces in early-stage lung adenocarcinoma: a retrospective study. *BMC Cancer.* 2021;21(1):913. doi:10.1186/s12885-021-08648-0
21. Chen X, Zhou H, Wu M, et al. Prognostic impact of spread through air spaces in patients with ≤2 cm stage IA lung adenocarcinoma. *J Thorac Dis.* 2024;16(4):2432-2442. doi:10.21037/jtd-24-444
22. Tian Y, Feng J, Jiang L, et al. Integration of clinicopathological and mutational data offers insight into lung cancer with tumor spread through air spaces. *Ann Transl Med.* 2021;9(12):985. doi:10.21037/atm-21-2256
23. Ye R, Yu Y, Zhao R, Han Y, Lu S. Comprehensive molecular characterizations of stage I–III lung adenocarcinoma with tumor spread through air spaces. *Front Genet.* 2023;14:1101443. doi:10.3389/fgene.2023.1101443
24. Kadota K, Nitadori JI, Sima CS, et al. Tumor spread through air spaces is an important pattern of invasion and impacts the frequency and location of recurrences after limited resection for small stage I lung adenocarcinomas. *J Thorac Oncol.* 2015;10(5):806-814. doi:10.1097/jto.0000000000000486
25. Toyokawa G, Yamada Y, Tagawa T, et al. Significance of spread through air spaces in resected pathological stage I lung adenocarcinoma. *Ann Thorac Surg.* 2018;105(6):1655-1663. doi:10.1016/j.athoracsur.2018.01.037
26. Kagimoto A, Tsutani Y, Kushitani K, et al. Segmentectomy vs lobectomy for clinical stage IA lung adenocarcinoma with spread through air spaces. *Ann Thorac Surg.* 2021;112(3):935-943. doi:10.1016/j.athoracsur.2020.09.020
27. Chen D, Wang X, Zhang F, et al. Could tumor spread through air spaces benefit from adjuvant chemotherapy in stage I lung adenocarcinoma? A multi-institutional study. *Ther Adv Med Oncol.* 2020;12:1758835920978147. doi:10.1177/1758835920978147

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj. 1705329

Anatolian Curr Med J. 2025;7(5):576-582

Differentiating AVP deficiency and primary polydipsia: a clinical and biochemical perspective

Beril Turan Erdoğan[1], Muzaffer Serdar Deniz[1], Belma Özlem Tural Balsak[1], Çağlar Keskin[1], Şefika Burçak Polat[2], Oya Topaloğlu[2], Reyhan Ersoy[2], Bekir Çakır[2]

[1]*Department of Endocrinology and Metabolism, Ankara Bilkent City Hospital, Ankara, Turkiye*

[2]*Division of Endocrinology and Metabolism, Department of Internal Medicine, Faculty of Medicine, Ankara Yıldırım Beyazıt University, Ankara, Turkiye*

Cite this article as: Turan Erdoğan B, Deniz MS, Tural Balsak BÖ, et al. Differentiating AVP deficiency and primary polydipsia: a clinical and biochemical perspective. *Anatolian Curr Med J.* 2025;7(5):576-582.

Received: 23.05.2025 ◆ **Accepted**: 14.08.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: The polyuria–polydipsia syndrome encompasses three major disorders—arginine vasopressin (AVP) deficiency, AVP resistance, and primary polydipsia. This study aimed to differentiate AVP deficiency from primary polydipsia by evaluating clinical features, biochemical markers, and anterior pituitary hormone levels, using the water deprivation test as the primary diagnostic modality.

Methods: This retrospective observational study included 34 adult patients with polyuria–polydipsia syndrome who underwent a standardized inpatient water deprivation test. Patients were categorized into AVP deficiency (complete or partial) or primary polydipsia based on urine osmolality responses to dehydration and desmopressin. Clinical data, daily fluid intake, and nocturia frequency were recorded. Serum electrolytes and anterior pituitary hormones (LH, GH) were analyzed.

Results: AVP deficiency was diagnosed in 76.4% of patients (58.8% complete, 17.6% partial), while 23.5% had primary polydipsia. LH and GH levels were significantly higher in the primary polydipsia group ($p=0.011$ and $p=0.028$, respectively), whereas AVP deficiency was associated with lower gonadotropin levels, especially in postoperative cases. Serum sodium, chloride, and magnesium levels were significantly lower in primary polydipsia ($p<0.05$), reflecting dilutional hyponatremia. Urine osmolality was significantly higher in primary polydipsia ($p=0.011$), indicating preserved concentrating ability. Nocturia occurred in 96.2% of patients with AVP deficiency versus 12.5% in primary polydipsia ($p<0.001$).

Conclusion: The water deprivation test remains a valuable diagnostic tool for differentiating AVP deficiency from primary polydipsia. These retrospective findings may serve as predictive indicators in the differential diagnosis, particularly in clinical settings where water deprivation tests and copeptin testing are not readily available. Incorporating nocturia frequency, serum electrolytes, and anterior pituitary hormone levels particularly LH and GH may improve diagnostic precision and facilitate individualized management.

Keywords: Polyuria–polydipsia syndrome, arginine vasopressin deficiency, primary polydipsia, water deprivation test

INTRODUCTION

The polyuria–polydipsia syndrome encompasses three major disorders characterized by excessive urine production (polyuria) and increased fluid intake (polydipsia): arginine vasopressin (AVP) deficiency (formerly central diabetes insipidus), AVP resistance (previously nephrogenic diabetes insipidus), and primary polydipsia.[1,2] Despite sharing similar clinical presentations, these conditions differ significantly in their pathophysiology and management.[3,4]

AVP deficiency results from impaired AVP synthesis or secretion in the hypothalamic–pituitary axis and typically necessitates treatment with desmopressin to maintain water homeostasis.[1,5] AVP resistance occurs when the renal collecting ducts do not respond to AVP, often due to hereditary receptor mutations or acquired causes such as lithium-induced nephrotoxicity.[6] Primary polydipsia, on the other hand, involves excessive water intake that suppresses endogenous AVP release; management generally consists of fluid restriction and, in some cases, psychological interventions.[2,7]

Although acute excessive water intake in primary polydipsia can lead to hyponatremia, most patients maintain normal sodium levels unless intake exceeds renal excretory capacity. In AVP deficiency or resistance, patients may develop hypernatremia and dehydration if fluid intake is inadequate, but many maintain normonatremia when water access is unrestricted.[6,7] Postoperative AVP deficiency, especially after neurosurgical or pituitary procedures, can compromise AVP secretion and disrupt the adjacent pituitary axis, resulting in

Corresponding Author: Beril Turan Erdoğan, berilturan@hotmail.com

diminished levels of anterior pituitary hormones such as LH and GH.[2] Psychiatric or stress-related factors may influence LH secretion, as stress has been shown to suppress LH pulsatility and reduce basal LH levels.[4,8] Hence, a comprehensive endocrine workup, including both biochemical and hormonal evaluations, is crucial to ensure accurate diagnosis and appropriate management.[4]

The water deprivation test is commonly used to differentiate these disorders by assessing the kidneys' capacity to concentrate urine during dehydration, followed by administration of exogenous desmopressin.[9] Although some centers continue to rely primarily on this test, newer techniques including hypertonic saline infusion and arginine-stimulated copeptin measurement have shown promise in diagnostic accuracy. Additionally, evaluating key serum electrolytes (sodium, chloride, magnesium) and anterior pituitary hormones (LH, GH, ACTH) may help distinguish among different types of polyuria–polydipsia syndromes.[3]

Despite the widespread use of the water deprivation test, differentiating between partial AVP deficiency and primary polydipsia remains challenging in clinical practice, especially in the absence of copeptin assays or when anterior pituitary function is not routinely evaluated. Furthermore, the integration of hormonal markers—such as LH and GH—into the diagnostic framework is not well established in most existing studies. Therefore, this study aims to not only validate traditional clinical and biochemical criteria but also to explore the potential diagnostic utility of anterior pituitary hormones in distinguishing AVP deficiency from primary polydipsia. By addressing these gaps, our 9 findings may enhance diagnostic precision and contribute to more tailored therapeutic strategies.

METHODS

This study has been approved by the Ethics Committee of Clinical Researches No. 1 at Ankara Bilkent City Hospital (Date: 06.09.2023, Decision No: E1-23-3947) and conducted in accordance with the Declaration of Helsinki.

This retrospective observational study included patients presenting with polyuria and polydipsia to the Endocrinology Clinic of Ankara Bilkent City Hospital from February 2019 to December 2024. A total of 35 patients underwent a standardized diagnostic evaluation, including the water deprivation test, conducted in the inpatient endocrinology unit.

Exclusion criteria included patients with significant comorbid conditions that could affect water metabolism or interfere with test interpretation. Specifically, patients were excluded if they had advanced chronic kidney disease (eGFR <30 ml/min/1.73 m^2), decompensated liver disease, uncontrolled diabetes mellitus, active malignancy, severe electrolyte disturbances at baseline (serum sodium <130 or >150 mEq/L), or inability to comply with test protocols due to cognitive or psychiatric disorders.

All patients underwent a standardized water deprivation test following hospital admission, in accordance with established clinical protocols.[10] Participants were instructed to abstain from alcohol, diuretics, and desmopressin use for 24 hours prior to testing. Baseline measurements of body weight, serum sodium concentration, and plasma osmolality were obtained at 8:00 a.m. Urine samples were collected and urine osmolality was measured hourly throughout the dehydration period. The test was terminated upon either a ≥10% increase in urine osmolality between two consecutive samples or a ≥2% reduction in body weight.

After test termination, 2 µg of intranasal desmopressin was administered. Urine osmolality was subsequently assessed at 30-minute intervals over the next two hours. A post-desmopressin increase in urine osmolality of ≥50% was considered consistent with complete AVP deficiency, while an increase between 10% and 50% indicated partial AVP deficiency. Minimal or no change in urine osmolality, in the context of an adequate baseline value (typically >500 mOsm/kg), was considered diagnostic of primary polydipsia.

A single patient met diagnostic criteria for AVP resistance during the water deprivation test and was subsequently excluded from the analysis. In this case, AVP resistance was suspected to be secondary to chronic lithium therapy

For analytical purposes, patients with complete and partial AVP deficiency were merged into a single group (AVP deficiency), due to overlapping clinical features and similar management strategies. This grouping also allowed for enhanced statistical power in between-group comparisons.

Demographic data (age, sex, BMI) and clinical characteristics (symptom duration in months) were recorded. Nocturia frequency was classified as fewer than three, three to five, or more than five occurrences per night. Daily fluid consumption and urine output were categorized into four groups: less than 5 liters, 5–10 liters, 10–15 liters, and more than 15 liters per day.

All participants underwent comprehensive biochemical and hormonal assessment during their initial outpatient endocrinology clinic visit.

All participants underwent comprehensive biochemical and hormonal assessment. Biochemical parameters included fasting blood glucose, blood urea nitrogen (BUN), creatinine, estimated glomerular filtration rate (eGFR), uric acid, albumin, serum electrolytes (sodium, potassium, chloride, calcium, magnesium, phosphorus), and plasma/urine osmolality. Urine specific gravity was also measured.

Hormonal evaluations involved thyroid hormones (TSH, free T4, free T3), gonadotropins (FSH, LH), sex hormones (estradiol, total testosterone), prolactin (PRL), adrenocorticotropic hormone (ACTH), cortisol, insulin-like growth factor-1 (IGF-1), and growth hormone (GH).

Statistical Analysis

All data analyses were conducted using IBM SPSS Statistics version 26.0. The normality of continuous variables was assessed using the Shapiro-Wilk test. Continuous variables with normal distribution were analyzed using the independent samples t-test, while those without normal distribution were compared using the Mann-Whitney U test. Categorical

variables were analyzed using Fisher's exact test. Descriptive statistics were expressed as mean±standard deviation (SD), median (minimum–maximum), or frequencies and percentages as appropriate. A p-value of <0.05 was considered statistically significant for all analyses.

RESULTS

Patient Characteristics

Of the initial 35 patients evaluated, one was excluded due to a diagnosis of AVP resistance based on test findings. The patient had a history of chronic lithium use, suggesting secondary nephrogenic diabetes insipidus.

A total of 34 patients were included in the final analysis. The mean age was 42.7±12.6 years, and 41.2% were male. The mean BMI was 30.5±5.3 kg/m². The average symptom duration was 91.97±155.3 months.

Based on the water deprivation test results, 58.8% (n=20) of patients were diagnosed with complete AVP deficiency, 17.6% (n=6) with partial AVP deficiency, and 23.5% (n=8) with primary polydipsia. Notably, none of the patients in the primary polydipsia group experienced significant weight loss during the dehydration phase, which further supported the diagnosis (Table 1).

Table 1. Demographic, clinical, and etiological characteristics of patients with polyuria and polydipsia (n=34; one patient with AVP resistance excluded)

	Patients n=34
Age (years), (SD)	42.7 (12.64)
Gender, female, n (%)	20 (58.8)
BMI (kg/m²) (SD)	30.5 (5.28)
Duration of polyuria and polydipsia (months) (SD)	91.97 (155.34)
Nocturia	
Yes, n (%)	26 (76.5)
No, n (%)	8 (23.5)
Frequency of nocturia	
<3, n (%)	12 (46.2)
3-5, n (%)	8 (30.8)
>5, n (%)	6 (23.1)
Fluid intake and output	
<5, n (%)	2 (5.9)
5-10, n (%)	20 (58.8)
10-15, n (%)	8 (23.5)
>15, n (%)	4 (11.8)
Diagnosis	
AVP deficiency, n (%)	20 (58.8)
Partial AVP deficiency, n (%)	6 (17.6)
Primary polydipsia, n (%)	8 (23.5)
Etiology	
Idiopathic, n (%)	13 (38.2)
Post-operative, n (%)	7 (20.6)
Psychogenic, n (%)	8 (23.5)
Hypophysitis, n (%)	4 (11.8)
Infundibulum metastasis, n (%)	1 (2.9)
Meningioma, n (%)	1 (2.9)

SD: Standard deviation, BMI: Body-mass index, AVP: Arginine vasopressin

Hormonal Parameters

Significant differences were observed in anterior pituitary hormone levels between groups. LH levels were significantly higher in the primary polydipsia group compared to the AVP deficiency group (11.78±11.99 vs. 5.65±7.77 U/L; p=0.011). Similarly, GH levels were elevated in primary polydipsia (0.43±0.39 µg/L vs. 0.27±0.54 µg/L; p=0.028).

ACTH levels were higher in the AVP deficiency group (31.90±13.70 pg/ml vs. 26.01±9.16 pg/ml), although this difference was not statistically significant (p=0.265). No significant differences were observed in cortisol, IGF-1, thyroid hormones, FSH, PRL, estradiol, or testosterone levels (Table 2).

Biochemical Parameters

Primary polydipsia patients demonstrated significantly lower serum sodium (138±4 vs. 142±3 mEq/L; p=0.030), chloride (105±4 vs. 108±3 mEq/L; p=0.040), and magnesium (1.8±0.1 vs. 2.0±0.2 mg/dl; p=0.010), reflecting dilutional effects from excessive fluid intake.

Urine osmolality was significantly higher in the primary polydipsia group (129±41 mOsm/kg) than in the AVP deficiency group (101±48 mOsm/kg; p=0.011. Plasma osmolality, creatinine, eGFR, calcium, and phosphorus levels did not differ significantly between the groups (Table 3).

Clinical Features

Nocturia was present in 76.5% of the cohort. It was significantly more common in the AVP deficiency group than in primary polydipsia (96.2% vs. 12.5%; p<0.001). All patients reporting more than five episodes of nocturia per night belonged to the AVP deficiency group.

Daily fluid intake ranged from <5 to >15 liters/day. Most patients (58.8%) reported intake between 5–10 liters. Although patients with primary polydipsia tended to consume more fluid than those with AVP deficiency, this difference did not reach statistical significance (p=0.322).

Regarding etiology, psychogenic causes were exclusively observed in the primary polydipsia group. In contrast, AVP deficiency was attributed to idiopathic causes in 26.9%, postoperative causes in 26.9%, neurohypophyseal signal loss in 17.6%, hypophysitis in 11.8%, and rare lesions such as infundibulum metastasis and meningioma in 2.9% each.

Among the postoperative cases (n=9), five had non-functioning pituitary adenomas and two had craniopharyngiomas. All seven developed anterior hypopituitarism after surgery. Additionally, both patients with pituitary stalk meningioma and metastatic infundibular involvement underwent neurosurgical intervention and subsequently developed postoperative anterior pituitary hormone deficiencies. No patient had preoperative hypopituitarism documented prior to intervention (Table 4).

DISCUSSION

This study evaluated the clinical, hormonal, and biochemical profiles of patients with AVP deficiency and primary polydipsia, using the water deprivation test as the principal diagnostic tool. Our findings reinforce the utility of this test for distinguishing between these two conditions based on urinary and hormonal responses. We acknowledge that our study design lacks a comparator based on a current gold standard such as copeptin-based diagnostics. As a result,

Table 2. Comparison of hormonal parameters between AVP deficiency and primary polydipsia groups

	Diagnosis								
	AVP deficiency (n=26)				Primary polydipsia (n=8)				p
	Mean	Min	Max	SD	Mean	Min	Max	SD	
TSH, mIU/L	2.127	0.008	4.750	1.338	1.881	0.010	4.300	1.248	0.648[a]
fT4, ng/dl	1.07	0.67	1.39	0.17	1.20	0.89	2.10	0.41	0.745[b]
fT3, ng/L	3.37	1.90	4.20	0.53	3.57	3.05	4.20	0.40	0.344[a]
FSH, U/L	9.9	0.3	57.6	13.1	17.9	3.4	78.4	25.8	0.626[b]
LH, U/L	5.65	0.07	31.20	7.77	11.78	3.50	39.60	11.99	0.011[b]
E2, U/L	39.3	11.8	205.0	39.2	102.0	11.8	370.0	118.9	0.149[b]
Total testosteron, µg/L	1.3427	0.0700	4.5900	1.6421	2.1763	0.0900	7.3700	2.7889	0.443[a]
PRL, µg/L	15.73	0.60	112.50	21.70	21.03	5.40	48.00	15.71	0.155[b]
ACTH, pg/ml	31.90	5.00	57.30	13.70	26.01	13.30	37.20	9.16	0.265[a]
Cortisol, µg/dl	13.9	3.7	24.4	4.6	17.3	7.7	23.9	5.1	0.087[a]
IGF-1, µg/L	98	15	261	50	132	89	220	47	0.094[a]
GH, µg/L	0.27	0.05	2.20	0.54	0.43	0.10	1.10	0.39	0.028[b]

AVP: Arginine vasopressin, Min: Minimum, Max: Maximum, SD: Standard deviation, TSH: Thyroid stimulating hormone, fT4: Free thyroxine, fT3: Free triiodothyronine, FSH: Follicle- stimulating hormone, LH: Luteinizing hormone, E2: Estrodiol, PRL: Prolactin, ACTH: Adrenocorticotropic hormone, IGF-1: Insulin-like growth factor 1, GH: Growth hormone, a: Independent Samples t-test, b: Mann-Whitney U test

Table 3. Comparison of biochemical parameters between AVP deficiency and primary polydipsia groups

	Diagnosis								
	AVP deficiency (n=26)				Primary polydipsia (n=8)				p
	Mean	Min	Max	SD	Mean	Min	Max	SD	
Duration of polyuria and polydipsia (month)	91	1	660	159	113	1	480	168	0.451[b]
Fasting blood glucose, mg/dl	90	70	110	10	86	74	108	10	0.238[b]
BUN, mg/dl	21	13	30	5	21	14	32	7	0.986[a]
Creatinine, mg/dl	0.75	0.56	1.06	0.15	0.68	0.58	0.87	0.10	0.249[a]
eGFR, ml/min/1.73m^2	104	74	134	16	115	96	132	11	0.215[b]
Uric asid, mg/dl	6.0	4.2	9.4	1.3	5.0	3.0	6.8	1.3	0.058[a]
Albumin, g/L	42.9	36.0	47.0	2.9	42.0	36.0	47.0	3.5	0.487[b]
Na, mEq/L	142	134	147	3	138	130	142	4	0.030[a]
K, mEq/L	4.2	3.6	5.1	0.3	4.2	3.9	4.7	0.3	0.982[a]
Cl, mEq/L	108	102	115	3	105	99	109	4	0.040[a]
Ca, mg/dl	9.5	8.8	10.3	0.4	9.4	8.5	10.4	0.6	0.341[a]
Mg, mg/dl	2.0	1.6	2.2	0.2	1.8	1.8	1.9	0.1	0.010[a]
P, mg/dl	4.0	2.1	5.0	0.6	4.3	3.4	5.0	0.6	0.314[a]
Plasma osmolarity, mOsm/kg	287	189	302	21	290	272	305	11	0.984[b]
Urine osmolarity, mOsm/kg	101	48	299	48	129	99	226	41	0.011[b]
Urine specific gravity	1004	1000	1009	2	1005	1002	1009	2	0.370[a]

AVP: Arginine vasopressin, Min: Minimum, Max: Maximum, SD: Standard deviation, BUN: Blood urea nitrite, eGFR: Estimated glomerular filtration rate, Na: Sodium, K: Potassium, Cl: Chlorine, Ca: Calcium, Mg: Magnesium, P: Phosphorus, a: Independent Samples t-test, b: Mann-Whitney U test

our findings reflect internal consistency rather than external validation. Nevertheless, in many clinical settings where copeptin assays are unavailable, the water deprivation test remains the most feasible diagnostic approach. Our study aimed to assess whether additional clinical and hormonal parameters could enhance the interpretative value of the WDT in such real-world contexts.

Although the water deprivation test has known limitations in differentiating AVP deficiency from primary polydipsia, it continues to be widely used due to its accessibility and simplicity. By integrating clinical indicators—such as nocturia frequency and fluid intake—with biochemical and hormonal findings, we aimed to strengthen the diagnostic utility of this conventional test in routine endocrinology practice.

In our cohort, AVP deficiency was the predominant diagnosis (58.8%), followed by partial AVP deficiency (17.6%) and primary polydipsia (23.5%). This distribution mirrors patterns reported in tertiary endocrinology centers, where a significant portion of patients with polyuria-polydipsia syndrome are ultimately classified as having a complete or partial AVP deficiency.[1,3] The mean age and BMI observed were consistent with previously described cohorts, reinforcing

Table 4. Comparison of nocturia, fluid intake, and etiological factors between AVP deficiency and primary polydipsia groups

		Diagnosis				
		AVP deficiency (n=26)	Primary polydipsia (n=8)	Total	x^2	p
Nocturia	Yes	25 (96.2%)	1 (12.5%)	26 (76.5%)	22.595	0.000[a]
	No	1 (3.8%)	7 (87.5%)	8 (23.5%)		
Frequency of nocturia	<3	11 (44.0%)	1 (100.0%)	12 (46.2%)	1.367	1.000[a]
	3-5	8 (32.0%)	0 (0.0%)	8 (30.8%)		
	>5	6 (24.0%)	0 (0.0%)	6 (23.1%)		
Fluid intake and output (litres)	<5	1 (3.8%)	1 (12.5%)	2 (5.9%)	3.146	0.322[a]
	5-10	17 (65.4%)	3 (37.5%)	20 (58.8%)		
	10-15	5 (19.2%)	3 (37.5%)	8 (23.5%)		
	15-20	3 (11.5%)	1 (12.5%)	4 (11.8%)		
Etiology	Idiopathic	13 (50.0%)	0 (0.0%)	13 (38.2%)	27.070	0.000[a]
	Post-operative	7 (27%)	0 (0.0%)	7 (20.6%)		
	Psychogenic	0 (0.0%)	8 (100.0%)	8 (23.5%)		
	Hypophysitis	4 (15.4%)	0 (0.0%)	4 (11.8%)		
	Infundibulum metastasis	1 (3.8%)	0 (0.0%)	1 (2.9%)		
	Menengiom	1 (3.8%)	0 (0.0%)	1 (2.9%)		

AVP: Arginine vasopressin, a: Fisher's Exact test

the generalizability of our sample.[4] The prolonged duration of symptoms observed in our cohort likely reflects a combination of under-recognition, delayed referral, and the intermittent or nonspecific nature of polyuria–polydipsia presentations. This pattern has been similarly reported in other studies and underscores the diagnostic challenges posed by these syndromes in real-world practice.

We observed significantly higher LH and GH levels in patients with primary polydipsia compared to those with AVP deficiency. These findings are consistent with the hypothesis that chronic psychogenic stress or psychiatric illness may stimulate hypothalamic-pituitary activation, leading to elevated gonadotropin and GH secretion.[2,4,9] Conversely, reduced levels of these hormones in AVP deficiency, particularly among postoperative patients, may reflect impaired hypothalamic input or direct surgical injury to the pituitary stalk.[2]

Although ACTH levels were not statistically different between groups, their tendency to be elevated in AVP deficiency aligns with prior observations of hypothalamic–pituitary–adrenal (HPA) axis compensation in the context of chronic free water loss.[5,11]

These findings suggest that selected anterior pituitary hormones—particularly LH and GH—may serve as adjunctive markers in the differential diagnosis of polyuria–polydipsia syndromes, especially in settings where copeptin measurement is not available.[8] GH secretion is known to exhibit significant interindividual variability and is influenced by multiple factors, including age, body composition, and comorbidities such as diabetes mellitus. While our study did not perform multivariate adjustment for these variables, the lack of significant difference in IGF-1 levels between groups supports the possibility that observed differences in GH may not reflect a true endocrine disturbance, but rather context-dependent variability.[12]

We considered menopausal status in the interpretation of LH values. While elevated LH levels in some primary polydipsia patients may reflect postmenopausal physiology, the overall trend observed—particularly in male patients and premenopausal women—suggests that additional factors such as psychogenic stress may also play a role.

The observed reduction in LH and GH levels in patients with AVP deficiency was particularly pronounced in those with a history of neurosurgical intervention. In our cohort, nine patients underwent surgery for sellar or suprasellar lesions, including pituitary adenomas, craniopharyngiomas, meningioma, and metastatic lesions. All of these patients developed anterior hypopituitarism postoperatively, whereas none had documented pituitary hormone deficiencies prior to surgery. These findings highlight the impact of surgical disruption of the hypothalamic–pituitary axis on anterior pituitary hormone secretion, particularly gonadotropins and GH. Therefore, when interpreting hormonal patterns especially in postoperative patients acquired pituitary insufficiency should be considered as a potential confounder in the differential diagnosis of polyuria–polydipsia syndromes.

Estradiol and testosterone levels are known to vary with age and reproductive status. Although we recorded these parameters, our analysis did not include age-adjusted stratification due to limited subgroup size. Importantly, these hormones were not central to the primary aim of the study and were analyzed for descriptive purposes. Future studies may incorporate age- and sex-specific hormone reference ranges to improve interpretability.

Patients with primary polydipsia exhibited significantly lower serum sodium, chloride, and magnesium levels, consistent with dilutional effects of excessive water consumption.[9,13] While hyponatremia may suggest primary polydipsia, particularly in acute presentations, it is not always sufficient to establish a definitive diagnosis. Some patients presented

with chronic, fluctuating symptoms or overlapping features, prompting further evaluation. In our cohort, water deprivation testing was conducted to clarify diagnosis and exclude other etiologies, especially in cases where clinical and biochemical findings were inconclusive at baseline. In contrast, the relative elevation of sodium and chloride levels in AVP deficiency reflects free water loss and the risk of hypernatremia if untreated.[1,7,14]

Interestingly, urine osmolality was significantly higher in the primary polydipsia group despite persistent polyuria. Although absolute urine osmolality values remained low in both groups, the relatively higher levels observed in primary polydipsia may reflect a partially preserved concentrating ability. This is consistent with previous studies suggesting that renal concentrating capacity may improve under supervised fluid restriction in chronic polydipsia cases. This suggests that at least partial concentrating ability is preserved in these patients and may become apparent during supervised fluid restriction. This observation supports previous reports indicating reversibility of renal concentrating dysfunction in primary polydipsia when excessive intake is curtailed.[11,14]

The absence of a significant difference in plasma osmolality between groups highlights a potential limitation of static plasma values in differentiating AVP deficiency from primary polydipsia, particularly at baseline.

Nocturia was found to be a highly discriminative symptom in our cohort, occurring in 96.2% of patients with AVP deficiency but only 12.5% of those with primary polydipsia. These findings align with the physiological role of AVP in promoting nocturnal antidiuresis and support prior reports that highlight nocturia as a clinical hallmark of AVP deficiency.[2,11] Moreover, all patients reporting >5 nocturnal episodes belonged to the AVP deficiency group, suggesting that frequent nocturia may serve as a clinically useful, albeit non-specific, diagnostic indicator.

Although daily fluid intake tended to be higher in primary polydipsia, this difference did not reach statistical significance—potentially due to overlapping behavior patterns and subjective variability in intake reporting.

From an etiological standpoint, our findings confirmed that psychogenic factors underlie all cases of primary polydipsia, while AVP deficiency was associated with a heterogeneous set of causes, including idiopathic, postoperative, inflammatory, and neoplastic conditions. These results underscore the need for comprehensive diagnostic evaluation including neuroimaging and clinical history in cases suggestive of AVP deficiency.

While the water deprivation test has limitations,[3] it remains accessible in most clinical settings. Our findings suggest that integrating clinical parameters such as nocturia frequency and anterior pituitary hormone levels may enhance diagnostic confidence in routine practice.

In recent years, the water deprivation test has been increasingly challenged by novel diagnostic methods such as copeptin-based testing. Among these, arginine-stimulated copeptin measurement has emerged as a promising alternative with notable advantages. Recent studies show that hypertonic saline-stimulated copeptin testing yields a sensitivity of 93.2% and a specificity of 100% in differentiating primary polydipsia from central diabetes insipidus, whereas the traditional water deprivation test offers an accuracy of only 73.3%.[8] Similarly, the arginine-stimulated test demonstrates a diagnostic accuracy of 93% (sensitivity 93%, specificity 92%) along with superior safety and patient comfort.[15] Unlike the water deprivation test, which requires prolonged inpatient monitoring (8–16 hours) and poses a risk of severe dehydration, copeptin assays can be performed in outpatient settings within 2–3 hours, with minimal side effects.[8,15] However, broader adoption of copeptin measurement remains limited due to its higher cost, requirement for specialized laboratory infrastructure, and restricted accessibility in many healthcare systems. In settings where copeptin testing is unavailable, our findings indicate that the diagnostic performance of the traditional water deprivation test may be improved by incorporating additional clinical indicators such as nocturia frequency and anterior pituitary hormone levels.

Limitations

We acknowledge the inherent limitation of relying solely on the WDT without external validation such as copeptin-based assays. Second, the relatively small sample size and unequal group distribution (26 patients with AVP deficiency vs. 8 with primary polydipsia) may limit the statistical power and generalizability of our findings. However, our study was not designed to re-validate the WDT itself, but rather to identify additional clinical and biochemical features that may aid interpretation in routine endocrinology settings where copeptin testing is unavailable. In this context, parameters such as nocturia frequency, serum electrolytes, and anterior pituitary hormone levels—particularly LH and GH—may improve diagnostic confidence when applied alongside the WDT.

Future research should focus on validating the diagnostic utility of anterior pituitary hormones particularly LH and GH as potential adjunctive markers in resource-limited settings. Moreover, large-scale, multicenter studies are needed to evaluate the long-term prognostic implications of these hormonal and biochemical patterns, and to refine diagnostic algorithms for polyuria–polydipsia syndrome.

CONCLUSION

In summary, this study highlights the continued relevance of the water deprivation test in differentiating AVP deficiency from primary polydipsia and suggests that selected anterior pituitary hormones may provide additional diagnostic insight. Clinical features such as nocturia frequency and biochemical markers including serum sodium and urine osmolality remain essential components of a comprehensive diagnostic approach. Tailoring diagnostic strategies to include endocrine and behavioral context may enhance accuracy and lead to more effective, individualized management plans.

ETHICAL DECLARATIONS

Ethics Committee Approval

This study has been approved by the Ethics Committee of Clinical Researches No. 1 at Ankara Bilkent City Hospital (Date: 06.09.2023, Decision No: E1-23-3947).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Refardt J, Winzeler B, Christ-Crain M. Diabetes insipidus: an update. *Endocrinol Metab Clin North Am*. 2020;49(3):517-531. doi:10.1016/j.ecl.2020.05.012
2. Sailer CO, Winzeler B, Christ-Crain M. Primary polydipsia in the medical and psychiatric patient: characteristics, complications and therapy. *Swiss Med Wkly*. 2017;147:w14514. doi:10.4414/smw.2017.14514
3. Teare H, Argente J, Dattani M, et al. Challenges and improvement needs in the care of patients with central diabetes insipidus. *Orphanet J Rare Dis*. 2022;17(1):58. doi:10.1186/s13023-022-02191-2
4. Atila C, Yıldız BO, Güler S, et al. Psychopathological characteristics in patients with arginine vasopressin deficiency (central diabetes insipidus) and primary polydipsia compared to healthy controls. *Eur J Endocrinol*. 2024;190(5):354-362. doi:10.1093/ejendo/lvae040
5. Itagaki E, Ozawa S, Yamaguchi S, et al. Increases in plasma ACTH and cortisol after hypertonic saline infusion in patients with central diabetes insipidus. *J Clin Endocrinol Metab*. 2001;86:5749-5754.
6. Flynn K, Thomas L, Jackson P, et al. Central and nephrogenic diabetes insipidus: updates on diagnosis and management. *Front Endocrinol (Lausanne)*. 2025;15:1479764. doi:10.3389/fendo.2024.1479764
7. Bologna K, Cesana-Nigro N, Refardt J, et al. Effect of arginine on the hypothalamic-pituitary-adrenal axis in individuals with and without vasopressin deficiency. *J Clin Endocrinol Metab*. 2020;105(3):e884-e892. doi:10.1210/clinem/dgaa157
8. Fenske W, Refardt J, Chifu I, et al. A copeptin-based approach in the diagnosis of diabetes insipidus. *N Engl J Med*. 2018;379(5):428-439. doi:10.1056/NEJMoa1803760
9. Pires BF, Andrade M, Silva T, et al. Primary polydipsia and sIDH Type D due to water-electrolytic disturbance in a schizophrenic patient: a case report and systematic review. *Open J Psychiatry*. 2025;15(1):1-15.
10. Pedrosa W, Drummond JB, Soares BS, Ribeiro-Oliveira A Jr. A combined outpatient and inpatient overnight water deprivation test is effective and safe in diagnosing patients with polyuria-polydipsia syndrome. *Endocr Pract*. 2018;24(11):963-972. doi:10.4158/ep-2018-0238
11. Pivonello R, Faggiano A, Filippella M, et al. Hypothalamus-pituitary-adrenal axis in central diabetes insipidus: ACTH and cortisol responsiveness to CRH administration. *J Endocrinol Invest*. 2002;25(10):932-937. doi:10.1007/BF03344064
12. Ho KK; 2007 GH Deficiency Consensus Workshop Participants. Consensus guidelines for the diagnosis and treatment of adults with GH deficiency II: a statement of the GH Research Society in association with the European Society for Pediatric Endocrinology, Lawson Wilkins Society, European Society of Endocrinology, Japan Endocrine Society, and Endocrine Society of Australia. *Eur J Endocrinol*. 2007;157(6):695-700. doi:10.1530/EJE-07-0631
13. Atila C, Lustenberger S, Yıldız BO, et al. Relationship between plasma urea and copeptin in response to arginine stimulation in healthy adults, patients with vasopressin deficiency, and primary polydipsia. *Pituitary*. 2025;28(1):18. doi:10.1007/s11102-024-01489-7
14. Trimpou P, Olsson DS, Ehn O, Ragnarsson O. Diagnostic value of the water deprivation test in the polyuria-polydipsia syndrome. *Hormones (Athens)*. 2017;16(4):414-422. doi:10.14310/horm.2002.1762
15. Winzeler B, Cesana-Nigro N, Refardt J, et al. Arginine-stimulated copeptin measurements in the differential diagnosis of diabetes insipidus: a prospective diagnostic study. *Lancet*. 2019;394(10198):587-595. doi:10.1016/S0140-6736(19)31255-3

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1752156

Anatolian Curr Med J. 2025;7(5):583-588

Comparison of thoracic epidural and peripheral regional analgesia techniques for postoperative pain management in video-assisted thoracoscopic surgery (VATS) procedures: a retrospective study

Ergün Mendeş, Doruk Yaylak, Yasemin Sincer, Mete Manici, Yavuz Gürkan

Department of Anesthesiology and Reanimation, Faculty of Medicine, Koç University, İstanbul, Turkiye

Cite this article as: Mendeş E, Yaylak D, Sincer Y, Manici M, Gürkan Y. Comparison of thoracic epidural and peripheral regional analgesia techniques for postoperative pain management in video-assisted thoracoscopic surgery (VATS) procedures: a retrospective study. *Anatolian Curr Med J.* 2025;7(5):583-588.

Received: 28.07.2025 • **Accepted**: 14.08.2025 • **Published**: 15.09.2025

ABSTRACT

Aims: Effective postoperative analgesia is essential in thoracic surgery to preserve respiratory function, facilitate early mobilization, and reduce pulmonary complications. Thoracic epidural analgesia (TEA) has traditionally been regarded as the gold standard for pain control. However, the increasing use of minimally invasive approaches such as video-assisted thoracoscopic surgery (VATS) has prompted the adoption of ultrasound-guided peripheral regional anesthesia techniques, including the paravertebral block (PVB), erector spinae plane block (ESPB), and serratus anterior plane block (SAPB). This study aimed to compare TEA and peripheral regional blocks in terms of postoperative pain scores, opioid consumption, and recovery parameters in patients undergoing VATS procedures.

Methods: This retrospective study included patients who underwent elective VATS between January 2023 and January 2025. Based on the analgesic technique received, patients were divided into two groups: TEA (n=82) and peripheral block group (PER, n=65). Pain was assessed using the Visual Analog Scale (VAS) at the post-anesthesia care unit (PACU) and at 24, 48, and 72 hours postoperatively. Opioid doses were standardized to morphine milligram equivalents (MME). Additional outcomes included intraoperative opioid use, postoperative rescue analgesia, length of hospital stay (LOS), complication rates, and 90-day mortality.

Results: Patients in the TEA group had significantly lower VAS scores at all time points compared to the PER group ($p<0.001$). PACU opioid administration and total MME consumption were also significantly reduced in the TEA group (0 mg [IQR 0–2] vs. 2 mg [IQR 0–4], $p<0.001$). Although intraoperative opioid use was generally similar, TEA provided greater opioid-sparing effects. No significant differences were observed between groups in LOS, complication rates, or 90-day mortality.

Conclusion: TEA demonstrated superior efficacy in postoperative pain control and opioid reduction compared to peripheral regional anesthesia techniques in patients undergoing VATS. While TEA remains the preferred option when feasible, ultrasound-guided peripheral blocks offer a valuable alternative, particularly in patients with contraindications to neuraxial techniques.

Keywords: Thoracic epidural analgesia, video-assisted thoracoscopic surgery, paravertebral block, erector spinae plane block, serratus anterior plane block, postoperative pain

INTRODUCTION

Postoperative pain following thoracic surgery remains a major clinical concern due to its negative impact on respiratory mechanics, delayed mobilization, and increased risk of pulmonary complications.[1] Effective pain control not only improves patient comfort but also contributes to preserving pulmonary function, minimizing complications, and reducing hospital length of stay (LOS).[2]

Thoracic epidural analgesia (TEA) has traditionally been considered the gold standard for postoperative pain management in thoracic procedures. By blocking nociceptive transmission at the spinal level, TEA provides effective and comprehensive analgesia.[3] However, its routine clinical use is often limited due to technical complexity, the risk of severe central complications, and contraindications in specific patient populations.[4,5]

In recent years, ultrasound-guided peripheral regional anesthesia techniques have gained significant attention as alternative approaches to TEA.[6] Among these, interfascial plane blocks such as the paravertebral block (PVB), erector spinae plane block (ESPB), and serratus anterior plane block (SAPB) have emerged as technically simpler, safer options associated with lower complication rates. PVB and ESPB are

Corresponding Author: Ergün Mendeş, erg.mendes@gmail.com

strongly recommended by the PROSPECT group after VATS, and SAPB is also among the techniques that can provide effective analgesia.[7] Although ESPB demonstrates superiority in intraoperative analgesia, it has also been shown that ESPB and SAPB provide similar efficacy in postoperative pain management.[8] The safety and efficacy of these blocks are increasingly supported by case reports and clinical trials.[9] As regional anesthesia approaches in thoracic surgery continue to evolve,[10] there remains a need for further comparative analyses between TEA and these newer techniques.

With the ongoing evolution of minimally invasive techniques, video-assisted thoracoscopic surgery (VATS) has become a preferred surgical approach due to reduced postoperative pain, shorter hospitalization, and fewer complications compared to open thoracotomy. The smaller incisions and limited tissue dissection in VATS procedures have encouraged the broader adoption of less invasive analgesic methods, particularly ultrasound-guided peripheral nerve blocks.[11]

Given the growing clinical experience and patient satisfaction associated with peripheral blocks, questions have emerged regarding their potential to replace TEA as the standard of care. In this context, the present retrospective study aimed to compare TEA and ultrasound-guided peripheral regional anesthesia techniques in terms of postoperative pain control, opioid consumption, and recovery outcomes in patients undergoing VATS. This study seeks to contribute to the ongoing optimization of analgesic strategies in thoracic surgery.

METHODS

Ethics

Ethical approval was obtained from the Koç University Committee on Human Researches (Date: 18.07.2025, Decision No: 2025.327.IRB1.054). The study was conducted in accordance with the principles outlined in the Declaration of Helsinki.

Study Design and Setting

This retrospective study was conducted using the shared clinical database of the Departments of Anesthesiology and Thoracic Surgery at Koç University Hospital. The study included patients who underwent VATS between January 2023 and January 2025.

Patient Selection

Patients were categorized into two groups based on the postoperative analgesic technique received: the TEA group, n=82 and the peripheral regional block group (PER group, n=65). Group allocation was determined according to standard clinical practice at the time of surgery, without randomization.

Inclusion criteria were as follows: age ≥18 years, elective VATS procedure (wedge resection, segmentectomy, or lobectomy), and availability of complete perioperative data. Exclusion criteria included thoracotomy, emergency surgery, incomplete medical records, reoperation within 72 hours, chronic opioid use prior to surgery, major intraoperative hemorrhage (>1.5 L), contraindications to regional anesthesia (e.g., coagulopathy or local infection), and procedures limited to diagnostic or minimal interventions (e.g., thoracentesis). Patients admitted to the intensive care unit (ICU) postoperatively were also excluded. Only those transferred to the ward after a standardized 1-hour follow-up in the post-anesthesia care unit (PACU) were included.

All VATS procedures were performed using a standardized two-port technique. The first (camera) port was placed in the 5th or 6th intercostal space along the mid-axillary line. The second (working) port was located in the 4th or 5th intercostal space, positioned either anteriorly or posteriorly depending on lesion location and surgical requirements. All operations were performed under general anesthesia using double-lumen endotracheal intubation to allow for single-lung ventilation.

Regional Anesthesia Techniques

All regional anesthesia procedures were performed under ultrasound guidance by an experienced anesthesiologist (M.M.), with over 10 years of experience in thoracic anesthesia and regional techniques. The block was administered at the end of surgery, immediately prior to extubation. Analgesic approaches included TEA, PVB, ESPB, or SAPB. The choice of technique was guided by surgical requirements and patient-specific anatomy, without a predetermined preference or bias. Patients in the TEA group received postoperative analgesia via epidural patient-controlled analgesia (PCA), while patients in the PER group received intravenous PCA.

Data Collection

Demographic data (age, height, weight), American Society of Anesthesiologists (ASA) physical status classification, and surgical parameters (type of VATS procedure, pleurodesis, mediastinal lymph node dissection [MLND], and operative time) were retrieved from electronic health records

Intraoperative opioid use was recorded and converted to morphine milligram equivalents (MME) to standardize comparisons. Postoperative opioid administration in the PACU was noted as present or absent, and total administered doses were also converted to MME (mg). Pain was assessed using the Visual Analog Scale (VAS; 0–10) at four time points: PACU, and 24, 48, and 72 hours postoperatively, based on nursing records. Other outcomes included LOS, postoperative complications, unplanned reintubation, and 90-day mortality.

Statistical Analysis

The data analyses were conducted using IBM SPSS Statistics version 27.0 (IBM Corp., Armonk, NY, USA). The Shapiro–Wilk test was used to assess the normality of distribution for continuous variables. Normally distributed variables were presented as mean±standard deviation (SD), and non-normally distributed variables were expressed as median and interquartile range (IQR; 25th–75th percentile). Between-group comparisons were made using the independent samples t-test or Mann–Whitney U test for continuous variables, and the Chi-square test or Fisher's exact test for categorical variables. A two-tailed p-value <0.05 was considered statistically significant.

RESULTS

A total of 415 thoracic surgical procedures were reviewed. After applying inclusion and exclusion criteria, 147 patients were included in the final analysis: 82 in the TEA group and 65 in the PER group (**Figure 1**).

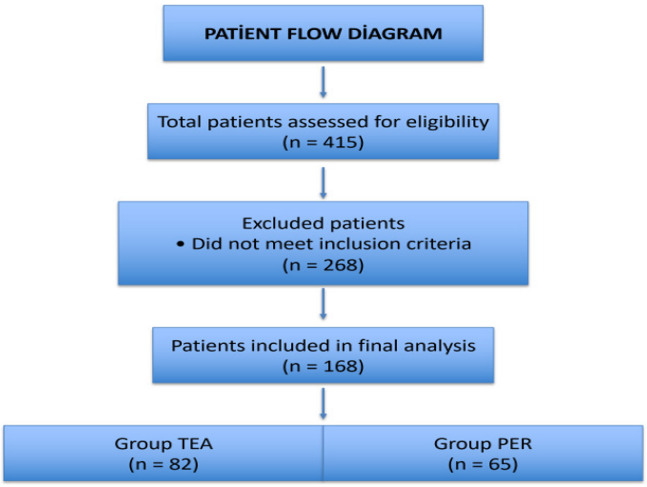


Figure 1. Patient flow diagram
TEA: Thoracic epidural analgesia, PER: Peripheral regional block

Demographic and Surgical Characteristics

The two groups were comparable in terms of baseline characteristics. The mean age was 62.41±13.53 years in the TEA group and 64.46±12.54 years in the PER group (p=0.344). No significant differences were observed in height (167.04±9.12 cm vs. 166.63±10.14 cm, p=0.802) or weight (73.37±14.76 kg vs. 74.35±13.16 kg, p=0.669). ASA physical status distribution (I/II/III) was similar between groups (TEA: 2/65/15 vs. PER: 1/57/7, p=0.401).

VATS performed—wedge resection, segmentectomy, or lobectomy—did not differ significantly between the groups (p=0.474). Pleurodesis was performed in four patients in the TEA group and two in the PER group (p=0.584). MLND was conducted in 54 of 82 patients in the TEA group and 38 of 65 in the PER group (p=0.358). The median operative time was 125 minutes (IQR 110–150) in the TEA group and 120 minutes (IQR 112.5–142.5) in the PER group, with no statistically significant difference (p=0.070).

Intraoperative and Postoperative Opioid Use

Intraoperative fentanyl (75 µg vs. 100 µg, p=0.735) and remifentanil (300 µg vs. 250 µg, p=0.994) doses were similar between groups. However, the use of other intraoperative opioids (converted to MME) was significantly lower in the TEA group (median 2 mg [IQR 2–2]) compared to the PER group (2 mg [IQR 2–3], p=0.000).

PACU, opioid use was significantly less frequent in the TEA group (32.9%) than in the PER group (70.8%) (p=0.000). Additionally, the median PACU opioid dose was significantly lower in the TEA group (0 mg [IQR 0–2]) than in the PER group (2 mg [IQR 0–4], p=0.000).

Postoperative Pain Scores

VAS scores were consistently lower in the TEA group at all time points:

PACU: TEA 2.00 [0.00–5.00] vs. PER 5.00 [3.00–6.00], p=0.000

24 hours: TEA 1.00 [0.00–3.00] vs. PER 3.00 [1.00–4.00], p=0.000

48 hours: TEA 0.00 [0.00–2.00] vs. PER 3.00 [1.00–4.00], p=0.000

72 hours: TEA 0.00 [0.00–1.00] vs. PER 2.00 [0.00–3.00], p=0.000 (**Figure 2**).

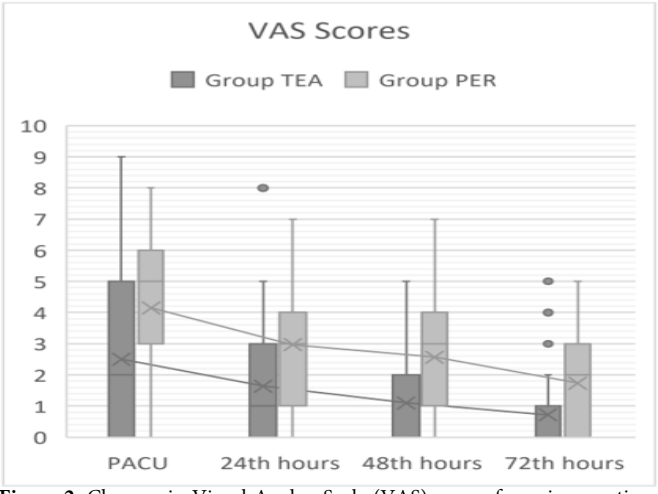


Figure 2. Changes in Visual Analog Scale (VAS) scores for pain over time, categorized by group
TEA: Thoracic epidural analgesia, PER: Peripheral regional block, PACU: Post-anesthesia care unit

Other Outcomes

LOS was 4 days in both groups (TEA: 4.00 [IQR 3–5] vs. PER: 4.00 [IQR 3–4], p=0.238). No patients in either group required unplanned reintubation or experienced weaning failure beyond 48 hours (p=1.000 for both). Ninety-day mortality was observed in one patient in the TEA group and two patients in the PER group (p=0.429). Overall postoperative complication rates were comparable (TEA: 7/82 [8.5%] vs. PER: 5/65 [7.7%], p=0.853) (**Table**).

DISCUSSION

This retrospective study compared TEA with ultrasound-guided peripheral regional anesthesia techniques in terms of postoperative pain control, opioid requirements, and recovery outcomes following VATS. The findings demonstrate that TEA provided significantly better postoperative analgesia than peripheral nerve blocks, as evidenced by consistently lower VAS scores and reduced opioid consumption at all time points up to 72 hours postoperatively.

In the TEA group, the median VAS score in the PACU was 2.00, compared to 5.00 in the peripheral block group (PER), with statistically significant differences maintained at 24, 48, and 72 hours. These findings are supported by the significantly lower need for rescue opioids in the PACU and lower total opioid consumption—expressed in MME—in the TEA group.

Table. Comparison of demographic characteristics, surgical data, analgesic requirements, and postoperative outcomes between group TEA and group PER patients undergoing video-assisted thoracoscopic surgery (VATS) procedures

	Group TEA (n=82)	Group PER (n=65)	p-value
Age (years)	62.41±13.53	64.46±12.54	0.344
Height (cm)	167.04±9.12	166.63±10.14	0.802
Weight (kg)	73.37±14.76	74.35±13.16	0.669
ASA scores I/II/III (n)	2/65/15	1/57/7	0.401
VATS wedge/segmentectomy/lobectomy (n)	41/34/7	38/24/3	0.474
Pleurodesis (-/+)	78/4	63/2	0.584
MLND (-/+)	28/54	27/38	0.358
Operative time (min)	125 (110-150)	120 (112.50-142.50)	0.070
intraoperative fentanyl (μg)	75 (50-100)	100 (50-100)	0.735
intraoperative remifentanil (μg)	300 (200-400)	250 (200-375)	0.994
Other opioid intraoperative (mg-MME)	20 (20-20)	20 (20-30)	0.000
PACU opioid use (-/+)	55/27	19/46	0.000
PACU opioid dose (mg-MME)	0 (0-20)	20 (0-40)	0.000
LOS (days)	4.00 (3-5)	4.00 (3-4)	0.238
VAS scores (PACU)	2.00 (0.00–5.00)	5 (3.00–6.00)	0.000
VAS scores (24th hours)	1.00 (0.00–3.00)	3.00 (1.00–4.00)	0.000
VAS scores (48th hours)	0.00 (0.00–2.00)	3.00 (1.00–4.00)	0.000
VAS scores (72th hours)	0.00 (0.00–1.00)	2.00 (0.00–3.00)	0.000
Complications (n)	7	5	0.853

Data presented as mean±standard derivation, median with interquartile range (25th–75th percentile) and n/n. TEA: Thoracic epidural analgesia, PER: Peripheral regional block, ASA: American Society of Anesthesiologists, VATS: Video-assisted thoracoscopic surgery, MLND: Mediastinal lymph node dissection, Min: Minimum, MME: Morphine milligram equivalent, PACU: Post-anesthesia care unit, LOS: Length of stay, VAS: Visual Analog Scale

Our results align with the findings of Adiyeke et al.,[12] who reported that although TEA and PVB yielded similar outcomes in terms of early pain scores and ICU admission, TEA demonstrated superiority in several secondary parameters. This advantage may be attributable to the broader analgesic coverage provided by TEA, which includes both parietal and visceral components of thoracic pain through blockade of sympathetic and visceral fibers.[13]

Among the peripheral techniques evaluated in this study, PVB is generally considered the most centrally acting, with some literature reporting analgesic efficacy comparable to that of TEA. However, unless contraindicated, TEA continues to be regarded as the gold standard for thoracic surgical analgesia due to its depth and duration of effect.[14] Similarly, TEA provides a broader range of analgesia by targeting both visceral and parietal pain components, giving it an advantage over peripheral techniques.[15] In our study, this clinical superiority was supported not only by subjective pain scores but also by objective measures, including reduced PACU opioid demand and lower analgesic consumption over the first three postoperative days.

In the study by Jo et al.,[16] it was found that all three procedures (PVB, ESPB and SAPB) reduced opioid use after VATS. PVB and ESPB were found to provide better pain control than SAPB. In contrast, a meta-analysis by Scorsese et al.[17] found no statistically significant advantage of TEA over peripheral blocks such as PVB, ESPB, and SAPB. Despite the lack of significance, TEA consistently achieved greater analgesic efficacy. These discrepancies in findings across studies may stem from variations in block techniques, catheter use (single-shot vs. continuous), local anesthetic volumes, and study design. Similarly, our study showed no significant differences between the groups in terms of LOS, or pulmonary complications. Notably, the use of epidural PCA in the TEA group versus IV PCA in the peripheral group may explain the observed superiority in VAS scores in the TEA group.

In a systematic review of 16 randomized controlled trials, Lin et al.[9] found that TEA, PVB and ESP provided effective postoperative analgesia after 24 hours, regardless of whether the single-shot or continuous catheter technique was used and excluding placebo or sham procedures. Despite the differences in PCA modality between the groups in our study, TEA showed superiority in both perioperative analgesic consumption and VAS scores up to 72 hours.

In our study, although TEA showed clear superiority in analgesic outcomes, there were no significant differences in LOS, complication rates, or 90-day mortality between the groups. This may be explained by the relatively low-risk nature of our patient, all of whom underwent minimally invasive VATS procedures and were managed postoperatively in the general ward.

It is important to emphasize that the inferior performance of peripheral blocks in our study should not imply ineffectiveness. Peripheral nerve blocks, especially ESPB and PVB, remain valuable options, particularly for patients with contraindications to neuraxial techniques or when TEA is technically challenging.[18] ESPB, in particular, is widely adopted due to its ease of use and low complication

rate, although its analgesic spread may be limited to parietal structures and may not adequately cover visceral pain.[19]

The significantly lower intraoperative opioid requirements in the TEA group further support the depth of analgesia provided by this technique. Effective intraoperative pain control has been associated with reduced postoperative opioid consumption and improved recovery trajectories. Our findings are consistent with previous studies demonstrating the widespread use and clinical effectiveness of regional anesthesia techniques in postoperative pain management after thoracic surgery. They are also in line with observations that TEA remains the most effective method following thoracotomy, while less invasive fascial plane blocks are preferred in VATS procedures.[20]

The variability in findings across the literature reflects the complex and multifactorial nature of regional anesthesia efficacy.[17,21] Factors such as patient anatomy, practitioner experience, type and timing of the block, and the local anesthetic regimen all influence clinical outcomes. In this context, recent clinical practice has reported the use of rhomboid intercostal plane block,[22] serratus superior posterior plane block,[23,24] the combined application of different regional techniques (e.g., PVB+ESPB),[25,26] or the use of the same block with different approaches (e.g., superficial+deep SAPB)[27] as additional options for enhancing postoperative analgesia after VATS. Furthermore, the extensive anatomical coverage of the recto-intercostal plane block—extending from the subxiphoid region to the lateral abdominal wall, as demonstrated in the cadaveric study by Tulgar et al.[28]—may help explain its potential clinical utility in thoracic and upper abdominal procedures. Future prospective studies with standardized protocols are needed to more clearly define the role of each technique in thoracic surgery.

Limitations

This study has several limitations that should be acknowledged. First, the PER group included a heterogeneous set of techniques—PVB, ESPB and SAPB—which differ anatomically, in mechanism of action, and in their potential for visceral spread. This heterogeneity limits the ability to make definitive conclusions about any single peripheral technique.

Second, the retrospective design of the study inherently carries risks of selection bias and unmeasured confounding variables. Although group allocation was based on standard clinical practice and not randomized, the possibility of differences in clinical decision-making that influenced the choice of analgesic technique cannot be excluded.

Third, different modes of postoperative analgesia were used across groups: epidural PCA in the TEA group and intravenous PCA in the PER group. This discrepancy may have affected pharmacokinetics and analgesic effectiveness, potentially biasing the comparison in favor of TEA.

Fourth, the study was conducted at a single academic center with a dedicated thoracic anesthesia team, which may limit generalizability to other clinical settings with different levels of expertise or resources.

Lastly, the study population consisted exclusively of relatively low-risk patients undergoing elective VATS. Therefore, the findings may not apply to patients undergoing thoracotomy, those at higher perioperative risk, or those requiring ICU management. Future prospective, multicenter randomized controlled trials focusing on standardized peripheral techniques and consistent analgesic protocols are necessary to validate and expand upon these findings.

CONCLUSION

This retrospective study demonstrated that TEA was superior to ultrasound-guided peripheral regional anesthesia techniques—specifically PVB, ESPB, and SAPB—in providing postoperative analgesia for patients undergoing VATS. TEA was associated with significantly lower VAS pain scores and reduced opioid requirements in both the intraoperative and early postoperative periods.

These findings support the continued use of TEA as the gold standard for postoperative pain management in thoracic surgery, particularly in patients who are eligible for neuraxial techniques and who may benefit from visceral as well as somatic analgesia. Nonetheless, ultrasound-guided peripheral nerve blocks remain valuable alternatives, offering effective and safe pain control, especially in cases where TEA is contraindicated or technically unfeasible.

As the use of minimally invasive surgical techniques continues to expand, further high-quality randomized controlled trials are warranted to clarify the comparative benefits of peripheral blocks and to optimize analgesic strategies tailored to individual patient profiles.

HIGHLIGHTS

- TEA provided significantly lower postoperative pain scores than peripheral regional techniques following VATS.
- TEA was associated with a substantial reduction in both intraoperative and postoperative opioid requirements, as measured by MME.
- Peripheral nerve blocks, including PVB, ESPB, and SAPB, were effective but less potent compared to TEA in managing postoperative pain.
- Despite differences in analgesic outcomes, LOS, complication rates, and 90-day mortality were similar between TEA and peripheral block groups.
- Ultrasound-guided peripheral blocks remain safe and viable alternatives, especially when TEA is contraindicated or technically challenging.

ETHICAL DECLARATIONS

Ethics Committee Approval

Ethical approval was obtained from the Koç University Committee on Human Researches (Date: 18.07.2025, Decision No: 2025.327.IRB1.054).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Makkad B, Heinke TL, Sheriffdeen R, et al. Practice advisory for postoperative pain management of thoracic surgical patients: a report from the society of cardiovascular anesthesiologists. *J Cardiothorac Vasc Anesth*. 2025;39(5):1306-1324. doi:10.1053/j.jvca.2024.12.004
2. Zhu J, Wei B, Wu L, et al. Effect of thoracic paravertebral block on postoperative pulmonary complications after video-assisted thoracoscopic surgery: a dual-center randomized clinical trial. Ther *Clin Risk Manag*. 2025;21:691-703. doi:10.2147/TCRM.S515093
3. Raft J, Richebé P. Anesthesia for thoracic ambulatory surgery. *Curr Opin Anaesthesiol*. 2019;32(6):735-742. doi:10.1097/ACO.0000000000000795
4. Hamilton C, Alfille P, Mountjoy J, Bao X. Regional anesthesia and acute perioperative pain management in thoracic surgery: a narrative review. *J Thorac Dis*. 2022;14(6):2276-2296. doi:10.21037/jtd-21-1740
5. Mendes E, Semerkant T, Cok OY. Application of a conventional paravertebral block on the thorax with a novel intrathoracic approach during the intraoperative period: two case reports: paravertebral block with an intrathoracic approach. *J Surg Med*. 2023;7(3):249-251. doi:10.28982/josam.1109449
6. Holm JH, Bak M, Brøchner AC. Epidural analgesia versus systemic opioids for postoperative pain management after VATS: protocol for a systematic review. *Acta Anaesthesiol Scand*. 2025;69(1):e14546. doi:10.1111/aas.14546
7. Feray S, Lubach J, Joshi GP, Bonnet F, Van de Velde M. PROSPECT guidelines for video-assisted thoracoscopic surgery: a systematic review and procedure-specific postoperative pain management recommendations. *Anaesthesia*. 2022;77(3):311-325. doi:10.1111/anae.15609
8. Demirci B, Akbudak İH, İlhan S, Mete Yıldız A. Comparison of erector spinae plane block and serratus anterior plane block for pain management in thoracoscopic surgery. *J Health Sci Med*. 2025;8(4):629-634. doi:10.32322/jhsm.1703176
9. Lin J, Liao Y, Gong C, et al. Regional analgesia in video-assisted thoracic surgery: a bayesian network meta-analysis. *Front Med (Lausanne)*. 2022;9:842332. doi:10.3389/fmed.2022.842332
10. Sandeep B, Huang X, Li Y, Xiong D, Zhu B, Xiao Z. A comparison of regional anesthesia techniques in patients undergoing video-assisted thoracic surgery: a network meta-analysis. *Int J Surg*. 2022;105:106840. doi:10.1016/j.ijsu.2022.106840
11. Piccioni F, Segat M, Falini S, et al. Enhanced recovery pathways in thoracic surgery from Italian VATS group: perioperative analgesia protocols. *J Thorac Dis*. 2018;10(Suppl 4):s555-s563. doi:10.21037/jtd.2017.12.86
12. Adıyeke Ö, Sarban O, Mendeş E, Abdullah T, Gümüş Özcan F. Comparison of the efficacy between thoracic epidural analgesia and paravertebral block in patients undergoing pneumonectomy: a retrospective cohort study. *GKD Anest Yoğ Bak Dern Derg*. 2024;30(1):29-34. doi:10.14744/GKDAD.2024.91668
13. Xiang Y, Chen L, Jia J, Yili F, Changwei W. The association of regional block with intraoperative opioid consumption in patients undergoing video-assisted thoracoscopic surgery: a single-center, retrospective study. *J Cardiothorac Surg*. 2024;19(1):124. doi:10.1186/s13019-024-02611-3
14. Dossi R, Patella M, Barozzi B, et al. Comparing erector spinae plane (ESP) and thoracic paravertebral (TPV) block analgesic effect after elective video-assisted thoracic surgery: a randomized, multiple-blinded, non-inferiority trial. *J Thorac Dis*. 2025;17(3):1531-1540. doi:10.21037/jtd-24-1548
15. Zeng J, Tang ZH, Liang JQ, et al. Comparison of various regional analgesia methods for postoperative analgesic effects in video-assisted thoracoscopic surgery: a systematic review and network meta-analysis. *Pain Physician*. 2022;25(7):E917-e30.
16. Jo Y, Park S, Oh C, et al. Regional analgesia techniques for video-assisted thoracic surgery: a frequentist network meta-analysis. *Korean J Anesthesiol*. 2022;75(3):231-244. doi:10.4097/kja.21330
17. Scorsese G, Jin Z, Greenspan S, et al. Effectiveness of thoracic wall blocks in video-assisted thoracoscopic surgery, a network meta-analysis. *J Pain Res*. 2023;16:707-724. doi:10.2147/JPR.S396530
18. Luo G, Tao J, Zhu J, Xie K, Ni C. Comparison of analgesic effects of different regional blocks in video-assisted thoracic and breast surgeries: a network meta-analysis and systematic review. *Pain Physician*. 2022;25(5):339-354.
19. Zhang S, Han X, Zhou D, et al. The effects of erector spinae plane block on perioperative opioid consumption and rehabilitation in video assisted thoracic surgery. *BMC Anesthesiol*. 2021;21(1):313. doi:10.1186/s12871-021-01536-x
20. Yilmaz M, Uzun N, Oral Ahiskalioglu E, Ulas A. Retrospective evaluation of analgesia approaches in patients undergoing thoracic surgery over the last one year: a single-center study. *Challenge J Perioperative Med*. 2025;3(1):1-7. doi:10.20528/cjpm.2025.01.001
21. Hong JM, Kim E, Jeon S, et al. A prospective double-blinded randomized control trial comparing erector spinae plane block to thoracic epidural analgesia for postoperative pain in video-assisted thoracic surgery. *Saudi Med J*. 2023;44(2):155-163. doi:10.15537/smj.2023.44.2.20220644
22. Ökmen K, Köprücüoğlu M. Rhomboid intercostal and serratus anterior interfascial plane blocks for the treatment of post-operative pain after video-assisted thoracoscopic surgery: a retrospective propensity-matched study. *Turk J Anaesthesiol Reanim*. 2021;49(3):211-217. doi:10.5152/TJAR.2020.471
23. Tulgar S, Ciftci B, Ahiskalioglu A, et al. Serratus posterior superior intercostal plane block: a technical report on the description of a novel periparavertebral block for thoracic pain. *Cureus*. 2023;15(2):e34582. doi:10.7759/cureus.34582
24. Manici M, Kalyoncu İ, Sincer Y, Gürkan Y. Enhancing minimally invasive thoracic surgery: efficacy of serratus posterior superior intercostal plane block in VATS procedures. *GKDA Derg*. 2024;30(2):87-88. doi:10.14744/GKDAD.2024.22448
25. Zhang L, Hu Y, Liu H, et al. Analgesic efficacy of combined thoracic paravertebral block and erector spinae plane block for video-assisted thoracic surgery: a prospective randomized clinical trial. *Med Sci Monit*. 2023;29:e940247. doi:10.12659/MSM.940247
26. Zengin M, Alagöz A, Sazak H, Ülger G, Baldemir R, Şentürk M. Comparison of efficacy of erector spinae plane block, thoracic paravertebral block, and erector spinae plane block and thoracic paravertebral block combination for acute pain after video-assisted thoracoscopic surgery: a randomized controlled study. *Minerva Anestesiol*. 2023;89(3):138-148. doi:10.23736/S0375-9393.22.16639-3
27. Zengin M, Baldemir R, Ülger G, Sazak H, Alagöz A. Comparison of deep and combined serratus anterior plane block after video-assisted thoracoscopic surgery; a prospective randomized trial. *J Health Sci Med*. 2023;6(1):18-24. doi:10.32322/jhsm.1185437
28. Tulgar S, Ciftci B, Ahiskalioglu A, et al. Recto-intercostal fascial plane block: another novel fascial plane block. *J Clin Anesth*. 2023;89:111163. doi:10.1016/j.jclinane.2023.111163

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1729053

Anatolian Curr Med J. 2025;7(5):589-596

Retrospective evaluation of mortality among stroke patients receiving thrombolytic therapy in the emergency department*

Emin Sancar[1], Abdurrahman Yılmaz[1], Gökçe Dağtekin[2], Eda Yaman[1], Sema Can[1]

[1]*Department of Emergency Medicine, Uşak Training and Research Hospital, Uşak, Turkiye*

[2]*Department of Public Health, Uşak Health Directorate, Uşak, Turkiye*

Cite this article as: Sancar E, Yılmaz A, Dağtekin G, Yaman E, Can S. Retrospective evaluation of mortality among stroke patients receiving thrombolytic therapy in the emergency department. *Anatolian Curr Med J.* 2025;7(5):589-596.

Received: 27.06.2025 ◆ **Accepted**: 17.08.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: Since strokes are an important cause of morbidity and mortality all over the world, early diagnosis and treatment are extremely important for sequela-free recovery. The best treatment option for an ischemic stroke is intravenous thrombolytic therapy with recombinant tissue plasminogen activator. In this study, we aimed to determine the factors affecting in-hospital mortality among patients who presented to the emergency department due to an ischemic stroke and received thrombolytic therapy.

Methods: This cross-sectional study reviewed the medical records of patients who presented to the emergency department with stroke symptoms between January 2018 and August 2022, were diagnosed with ischemic stroke, received thrombolytic treatment within the first 4.5 hours of symptom onset, and were subsequently hospitalized.

Results: After excluding patients who did not meet the inclusion criteria, 184 patients were included in the study. It was found that 42 (22.8%) of the patients had an exitus in the follow-up. We determined that the mortality rate was higher in patients admitted to the emergency department with emergency health services compared to outpatients ($p<0.05$). Mortality frequency was higher in patients who had a National Institutes of Health Stroke Scale (NIHSS) score ≥21 or were hospitalized for more than 14 days ($p<0.05$). In univariate analysis, age, admission route, length of hospital stay, and NIHSS score were associated with mortality. Multiple logistic regression analysis demonstrated that patients with an NIHSS score ≥21 had an 18.49-fold higher risk of death (95% CI: 6.55–52.20) compared to those with a score ≤20.

Conclusion: In our study, it was observed that the mortality rate was increased in patients who were aged 65 and over, had diabetes mellitus or atrial fibrillation comorbidities, were admitted to the hospital with emergency health services, were hospitalized for more than 14 days, and had NIHSS score of 21 and above.

Keywords: Ischemic stroke, thrombolytic therapy, mortality

*The present data of this study were presented orally at the 20th National Emergency Medicine Congress & 11th Intercontinental Emergency Medicine Congress & 11th International Critical Care and Emergency Medicine Congress held in Antalya on 16-19 May 2024. The abstracts were published in the Congress book.

INTRODUCTION

Stroke is a clinical condition characterized by the sudden onset of unilateral/bilateral weakness, numbness, visual field defects, diplopia, anisocoria, aphasia, ataxia, or other neurological deficits, resulting from infarction or hemorrhage in one or more parts of the nervous system.

According to the World Health Organization (WHO), approximately 14 million new stroke cases occur annually, leading to 5.5 million deaths.[1] Stroke is the second leading cause of mortality and the third leading cause of morbidity worldwide.[1] In a 2020 report published by the American Heart Association (AHA), stroke ranked as the fifth leading cause of death in the United States.[2] Globally, about one in four adults will experience a stroke during their lifetime, with a higher incidence in males compared to females.[1,3]

The etiology of stroke can be broadly classified as ischemic or hemorrhagic. Approximately 88% of all strokes are ischemic, 10% are intracerebral hemorrhages (ICH), and 2% are subarachnoid hemorrhages (SAH).[4] Ischemic strokes (IS) are caused by pathologies such as a subtotal arterial occlusion, which reduces cerebral blood flow; total occlusion, which completely halts cerebral perfusion; or systemic hypoperfusion.[5]

The penumbra refers to the potentially salvageable area of the brain that has experienced ischemia during an acute ischemic stroke (AIS) but has not yet undergone necrosis. It is a damaged but functionally viable region that can receive oxygen and nutrients through collateral circulation despite reduced blood flow. Timely reperfusion can preserve the penumbra and

Corresponding Author: Eda Yaman, edayamancbu2017@hotmail.com

prevent permanent neuronal injury.[6] Therefore, the primary goal of AIS treatment is to salvage this region.

It has been shown that there is a significant increase in recovery without sequelae with effective treatment options such as intravenous thrombolytic (IVT) and endovascular thrombectomy.[7] The most effective treatment recommended for patients diagnosed with AIS within the first 4.5 hours of symptom onset is intravenous recombinant tissue plasminogen activator (rtPA).[8] Alteplase, tenecteplase, and reteplase are rtPA.[9] Clinical studies regarding the use of tenecteplase and reteplase are ongoing, which is why their use in AIS treatment remains controversial today.[9,10] Patients undergoing IVT treatment should be regularly assessed for systemic bleeding, and neurological examination should be repeated to evaluate for ICH.

The primary objective of this study is to evaluate factors influencing in-hospital mortality in AIS patients who present to the emergency department (ED) and receive IVT treatment. The secondary objective is to assess the complications that occur in these patients and their associated conditions.

METHODS

Ethics

This retrospective cross-sectional study was conducted at Uşak Research and Training Hospital. The study protocol was approved by the Uşak University Faculty of Medicine Non-interventional Clinical Researches Ethics Committee (Date: 02.02.2023, Decision No: 65-65-19). The study was carried out in accordance with the principles of the Declaration of Helsinki. Given its retrospective design, the requirement for written informed consent was waived.

Study Design and Population

Uşak Research and Training Hospital is the only tertiary hospital in Uşak province. The ED serves approximately 1.000–1.500 patients daily. Between 01.01.2018 and 31.08.2022, patients being over 18 years old who presented to the Uşak Research and Training Hospital ED with stroke symptoms (sudden weakness, aphasia, dysarthria, diplopia, ataxia, altered consciousness, etc.) and were diagnosed with stroke were retrospectively screened.

Exclusion criteria

Patients were excluded if they met any of the following:

- Did not receive IVT therapy
- Missing essential clinical data
- Transferred to another institution
- Presentation>4.5 hours after symptom onset
- Unknown symptom onset time
- Stroke recognized upon awakening
- Presence of systemic malignancy
- Major surgery within the previous 3 weeks
- Gastrointestinal or genitourinary bleeding within the previous 3 weeks
- History of ischemic stroke within the previous 3 months
- Cranial or spinal trauma/surgery within the previous 3 months

The application of the exclusion criteria and the number of patients included in the study are summarized in **Figure**.

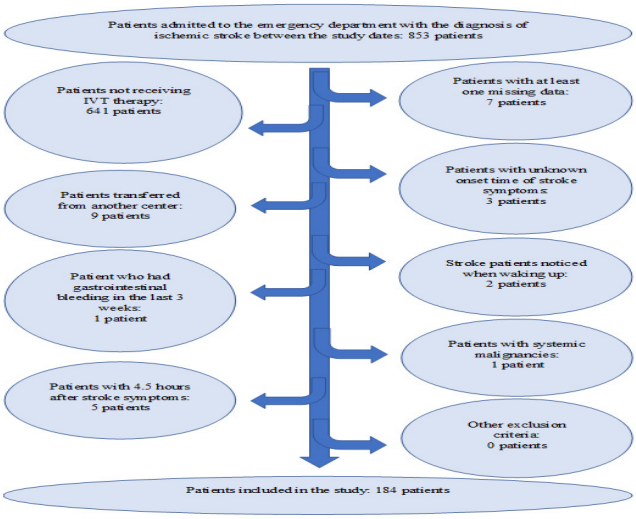


Figure. Exclusion criteria and patient numbers

Data Collection

Medical records of all included patients were reviewed from electronic databases and patient files in the hospital archives. The following information was assessed for all patients: age, gender, comorbidities, mode of hospital admission (outpatient, emergency health services (EHS)), laboratory results (complete blood count, renal function tests, electrolytes, cardiac markers, coagulation parameters), length of hospital stay (LOHS), complications, National Institutes of Health Stroke Scale (NIHSS) score, and in-hospital mortality (IHM).

NIHSS Score

The NIHSS is a universally accepted and highly reliable stroke scale used for prognosis determination and evaluating the short-term effectiveness of IVT and intra-arterial treatments in AIS.[11] In our study, patients were categorized into four groups based on their NIHSS scores: minor stroke (1-4), moderate stroke (5-15), moderately severe stroke (16-20), and severe stroke (21-42).

Thrombolytic Treatment Protocol

Not all AIS patients are eligible for IVT treatment. IVT treatment is not administered to patients whose symptom onset time exceeds 4.5 hours, who have evidence of acute bleeding on imaging, who have extensive hypodensity on computed tomography, who have a platelet count below 100.000/mm^3, whose INR value is above 1.7, whose aPTT value is longer than 40 seconds, or who have systolic blood pressure higher than 185 mmHg or diastolic blood pressure higher than 110 mmHg. In our hospital, for conscious and competent patients

who are planned to receive IVT, informed consent is obtained by the ED physician and neurology specialist after explaining the benefits and potential complications of IVT treatment. In cases where the patient lacks decision-making capacity and competence, informed consent is obtained from their primary caregivers using the same procedure. For those who refuse the treatment, informed refusal consent is obtained after ensuring that they have been adequately informed. For patients who provide consent, alteplase administration is calculated based on the recommended dosage of 0.9 mg/kg, with 10% of the calculated dose given as an intravenous bolus, followed by the remaining amount administered as a one-hour intravenous infusion, as described in the Turkish Ministry of Health AIS diagnosis and treatment guidelines.[12]

Statistical Analysis

The data analyses were conducted using SPSS (Statistical Package for the Social Sciences) version 15 (SPSS Inc.; Chicago, IL, USA). The normality of the data distribution was examined using the Kolmogorov-Smirnov and Shapiro-Wilk tests. Descriptive statistics were presented as numbers and percentages for categorical variables, and as means and standard deviations or minimum-maximum values for continuous variables. The distribution of categorical variables in cross-tabulations was compared using the Chi-square test. The t-test or Mann-Whitney U test was used for the comparison of continuous variables between two independent groups. A significance level of $p<0.05$ was considered statistically significant. To identify independent predictors of mortality, multivariate logistic regression analysis was conducted. A p-value of less than 0.05 was considered statistically significant for all analyses.

RESULTS

Out of the 184 patients included in the study, 104 (56.5%) were male, while 80 (43.5%) were female. The age of the patients ranged from 33 to 94, with a mean age of 71.5±12.2. Among the patients, 52 (28.3%) were below the age of 65, while 132 (71.7%) were 65 years and older. Of the patients, 62 (33.7%) arrived at the ED on foot, while 122 (66.3%) were brought in by EHS.

Patients whose International Classification of Diseases (ICD) diagnosis code was previously entered in the hospital system for diseases such as DM, HTN, AF, coronary artery disease (CAD), hyperlipidemia, previous CVD, congestive heart failure (CHF), and patients for whom prescription treatments were issued for these diseases were considered as patients with comorbidities. The accompanying comorbidities of the patients were as follows: HTN (n=180), AF (n=82), CAD (n=62), DM (n=59), hyperlipidemia (n=22), previous CVD (n=20), and CHF (n=16).

The LOHS for the patients ranged from 1.0 to 155.0 days, with a mean±SD of 13.8±17.4 days. The NIHSS scores of the patients ranged from 3.0 to 29.0, with a mean±SD of 13.5±6.5. Two patients (1.1%) in our patient group were classified as having a minor stroke. It was determined that 117 patients (63.6%) had a moderate stroke based on their NIHSS scores. 26 patients (14.1%) had a moderately severe stroke, and 39 patients (21.2%) had a severe stroke. A weak positive correlation was found between the NIHSS scores and the LOHS in the study group (r: 0.365, $p<0.001$).

It was found that 142 patients (77.2%) were discharged, while 42 patients (22.8%) had an IHM. There was no significant difference in terms of mortality between male and female genders, but it was higher in patients aged 65 and above and those who were brought to the hospital by EHS (p=0.538, p=0.007, $p<0.001$). It was observed that patients with a severe stroke had a higher mortality rate ($p<0.001$). Patients who stayed in the hospital for more than 14 days had a higher mortality rate (p=0.022). In terms of chronic diseases in the study group, although HTN was the most common risk factor, there was no difference in terms of mortality between patients with and without HTN (p=0.917). It was determined that mortality rate was higher in patients with AF and DM (p=0.01, p=0.014). The sociodemographic and clinical characteristics of the patients and their relationship with mortality are shown in **Table 1**.

Table 1. The relationship between sociodemographic and clinical characteristics of the cases and mortality

Variables	Survivors n (%)	Non-survivors n (%)	x^2; p
Gender			
Male	82 (78.8)	22 (21.2)	0.380; 0.538
Female	60 (75)	20 (25)	
Age			
<65 years	47 (90.4)	5 (9.6)	7.181; 0.007
≥65 years	95 (72)	37 (28)	
Form of arrival			
Ambulatory	59 (95.2)	3 (4.8)	17.174; <0.001
Emergency health services	83 (68)	39 (32)	
Length of stay in hospital			
14 days and less	107 (81.7)	24 (18.3)	5.241; 0.022
15 days and more	35 (66)	18 (34)	
NIHSS score			
20 and below	131(90.3)	14 (9.7)	67.368; <0.001
21 and over	11 (28.2)	28 (71.8)	
Hypertension			
-	3 (75.0)	1 (25.0)	0.011; 0.917
+	139 (77.2)	41 (22.8)	
Diabetes mellitus			
-	103 (82.4)	22 (17.6)	6.044; 0.014
+	39 (66.1)	20 (33.9)	
Atrial fibrillation			
-	86 (84.3)	16 (15.7)	6.628; 0.010
+	56 (68.3)	26 (31.7)	
Coronary artery disease			
-	94 (77.0)	28 (23.0)	0.003; 0.955
+	48 (77.4)	14 (22.6)	
Heart failure			
-	130 (77.4)	38 (22.6)	0.047; 0.828
+	12 (75.0)	4 (25.0)	
Previous CVD			
-	126 (76.8)	38 (23.2)	0.102; 0.750
+	16 (80.0)	4 (20.0)	
Hyperlipidemia			
-	123 (75.9)	39 (24.1)	1.198; 0.274
+	19 (86.4)	3 (13.6)	
Total	142 (77.2)	42 (22.8)	184 (100.0)

NIHSS: National Institutes of Health Stroke Scale, CVD: Cerebrovascular disease

A comparison was made between the group of patients who survived and the group who had an exitus using the Mann-Whitney U test for these blood parameters. A significant difference was found between the two groups for sodium and blood urea nitrogen (BUN) values. The distribution of laboratory findings according to the mortality status of the cases is listed in **Table 2**.

In the study, only the presence of bleeding complications was evaluated as a complication and the frequency of experiencing complications was found to be 20.7% (n=38). Cerebral hemorrhage was observed in 37 patients, while gastrointestinal bleeding was detected in 1 patient. It was determined that the frequency of complications increased in patients who applied with EHS as opposed to outpatient admission ($p<0.05$). It was also determined that the frequency of complications increased in the patient group with a hospital stay longer than 14 days and in the patient group with an NIHSS score greater than 21 ($p<0.05$). There was no significant association between comorbidities and the frequency of complications. The distribution of sociodemographic and clinical characteristics of patients with regards to the development of complications is shown in **Table 3**.

In the multivariate logistic regression analysis conducted with age, arrival status, LOHS, and NIHSS score, which were considered to be associated with mortality based on the results of univariate analysis, patients with an NIHSS score of 21 or higher were found to be 18.49 times (95% CI: 6.55-52.20) more at risk of mortality compared to those with a score of 20 or lower. The results of the multivariate logistic regression analysis showing the risk factors for mortality are presented in **Table 4**.

In the multivariate logistic regression analysis conducted with arrival status, LOHS, and NIHSS score, which were considered to be associated with the occurrence of complications based on the results of univariate analysis, patients with a hospital stay of 15 days or more were found to be 2.64 times (95% CI: 1.19-5.87) more at risk of developing complications compared to those with a hospital stay of 14 days or less ($p=0.017$). Patients admitted through EHS were found to be 1.88 (0.67-5.24) OR (95% CI) times more at risk of complications compared to ambulatory patients ($p=0.226$). Patients with an NIHSS score of 21 or higher were found to be 2.69 times (95% CI: 1.11-6.52) more at risk of developing complications compared to those with an NIHSS score of 20 or lower ($p=0.028$).

DISCUSSION

In our study, it was observed that the majority of patients who experienced IS were over the age of 65 and male, which is consistent with the literature.[13] However, the studies conducted by Ji et al.[14] and Thrift et al.[15] did not find a significant association between mortality risk and gender. Similarly, our study revealed no difference in mortality or complication risk between males and females. Mortality was higher among patients aged ≥65 years, while complication frequency was not significantly affected. This finding aligns with the results reported by Thrift et al.[15] The increased mortality risk in IS patients over 65 years may be explained by the higher prevalence of comorbidities in older individuals and the role of these comorbidities in increasing IS risk.

In our study, the most common comorbidities accompanying stroke were HTN, AF, CAD, and DM. When examining the literature, it is evident that HTN is the most common

Table 2. Distribution of laboratory findings according to mortality status of the cases

Variables	Survivors n=142		Non-survivors n=42		p value
	Mean±SD	Min-max	Mean±SD	Min-max	
Glucose (mg/dl)	151.88±71.06	64-552	163.1±61.66	83-343	0.069*
APTT (sec)	22.34±3.4	14.5-44	22.07±2.91	15.4-27.6	0.807*
Neutrophil lymphocyte ratio	3.51±3.56	0.5-33.63	4.09±6.34	0.7-38.15	0.163*
Creatinine (mg/dl)	1±0.41	0.6-4.3	1.04±0.33	0.66-2.18	0.229*
BUN (mg/dl)	19.33±8.01	9-75	23.19±8.21	12-60	0.001*
Sodium (mmol/L)	139.35±2.54	132-147	140.26±2.95	133-145	0.019*
Potassium (mmol/L)	4.22±0.47	3.2-5.5	4.36±0.44	3.3-5.4	0.057*
Calcium (mg/dl)	8.89±0.53	7.7-10.1	8.8±0.46	7.7-9.6	0.356*
C-reactive protein (mg/L)	8.79±19.58	0-167.6	11.98±27.32	0-129.6	0.53*
Hemoglobin (g/dl)	13.17±1.9	7.3-17.2	13.16±1.77	9.2-15.8	0.957*
Platelet (× 10^3/mcL)	254.13±82.23	100-573	229.29±68.9	126-403	0.064*
Leukocyte (× 10^3/mcL)	8.48±2.58	4-19.35	9.27±2.82	5.3-15.81	0.101*
Neutrophil (× 10^3/mcL)	5.62±2.38	2.19-15.69	5.91±2.64	2.41-13.52	0.673*
Lymphocyte (× 10^3/mcL)	2.11±1.02	0.24-7.65	2.51±1.32	0.34-5.99	0.119*
Eosinophil (× 10^3/mcL)	0.18±0.16	0-0.96	0.22±0.18	0-0.71	0.223*
Basophil (× 10^3/mcL)	0.04±0.05	0-0.5	0.04±0.02	0-0.09	0.973*
Monocytes (× 10^3/mcL)	0.53±0.22	0.2-1.9	0.6±0.26	0.25-1.32	0.221*
Potassium glucose ratio	0.03±0.01	0.01-0.05	0.03±0.01	0.01-0.05	0.259*

*Independent-Samples Mann-Whitney U test, SD: Standard deviation, Min: Minimum, Max: Maximum, BUN: Blood urea nitrogen

Table 3. Distribution of the cases in terms of sociodemographic and clinical characteristics and the frequency of complication development

Variables	Complication (-) n (%)	Complication (+) n (%)	x^2; p
Gender			
Male	82 (78.8)	22 (21.2)	0.037; 0.848
Female	64 (80.0)	16 (20.0)	
Age			
<65 years	44 (84.6)	8 (15.4)	1.227; 0.268
≥65 years	102 (77.3)	30 (22.7)	
Form of arrival			
Ambulatory	56 (90.3)	6 (9.7)	6.873; 0.009
Emergency health services	90 (73.8)	32 (26.2)	
Length of stay in hospital			
14 days and less	113 (86.3)	18 (13.7)	13.258; <0.001
15 days and more	33 (62.3)	20 (37.7)	
NIHSS score			
20 and below	124 (85.5)	21 (14.5)	15.889; <0.001
21 and over	22 (56.4)	17 (43.6)	
Hypertension			
-	4 (100.0)	0 (0.0)	1.064; 0.302
+	142 (78.9)	38 (21.1)	
Diabetes mellitus			
-	98 (78.4)	27 (21.6)	0.214; 0.644
+	48 (81.4)	11 (18.6)	
Atrial fibrillation			
-	83 (81.4)	19 (18.6)	0.573; 0.449
+	63 (76.8)	19 (23.2)	
Coronary artery disease			
-	99 (81.1)	23 (18.9)	0.716; 0.398
+	47 (75.8)	15 (24.2)	
Heart failure			
-	136 (81.0)	32 (19.0)	3.035; 0.081
+	10 (62.5)	6 (37.5)	
Previous CVD			
-	131 (79.9)	33 (20.1)	0.259; 0.611
+	15 (75.0)	5 (25.0)	
Hyperlipidemia			
-	126 (77.8)	36 (22.2)	2.038; 0.153
+	20 (90.9)	2 (9.1)	
Total	146 (79.3)	38 (20.7)	184 (100.0)

NIHSS: National Institutes of Health Stroke Scale, CVD: Cerebrovascular disease

Table 4. Multivariate logistic regression analysis showing mortality risk factors

Variables	Odds ratio	95% CI	p
Age Reference: 64 and below	2.83	0.85-9.40	0.088
Form of arrival Reference: EHS	2.65	0.69-10.08	0.153
Length of stay Reference: 14 days and less	0.74	0.27-2.05	0.567
NIHSS Reference: 20 and below points	18.49	6.55-52.20	<0.001
Constant: B=-3.645; p<0.001			

EHS: Emergency health services, NIHSS: National Institutes of Health Stroke Scale

comorbidity associated with IS.[3,16] The reason for the frequent occurrence of HTN comorbidity in IS patients in our study may be the irregular living and nutrition of the human population living in Turkiye. People living in rural areas visit health institutions less often and have fewer health checks. Previous studies in Turkiye have reported HTN prevalence rates in IS patients ranging from 58.8% to 72.5%.[17-20] Further research is needed to investigate the underlying causes of this high prevalence in our region.

While no significant relationship was found between HTN and mortality frequency in our study, increased mortality has been associated with HTN in the literature.[21] This could be attributed to the small sample size of our study. Moreover, while we were able to determine HTN history from patient records, data regarding antihypertensive medication use and lifestyle modifications were unavailable, making it unclear whether patients had controlled blood pressure. Due to incomplete medical records, reliable assessment of blood pressure at presentation was also not possible. These factors might have influenced the lack of a significant relationship between HTN and mortality in our study.

We found that patients with comorbidities such as DM and AF had a higher frequency of mortality. In a meta-analysis published by Lau et al.,[22] the prevalence of DM in patients with IS was found to be 33%. In our study, the prevalence was determined to be 32.1%. Obtaining a value close to the meta-analysis, which reviewed 66 articles, increases the reliability of our study. DM has been found to be associated with increased mortality, longer hospitalization periods, and recurrent admissions and strokes. Both our study and the literature show similar results, indicating that DM increases mortality. This may be attributed to the fact that DM affects multiple organs and shortens the expected lifespan, even in the absence of a stroke. Additionally, it is known that DM delays or impedes the healing of damaged tissues, including cerebral tissue in patients with IS. Some stroke patients become bedridden and develop pressure ulcers due to their inability to perform activities of daily living. The delayed healing of these pressure ulcers in DM patients may lead to infectious complications and severe conditions such as sepsis, which could increase the risk of mortality.

Similarly, AF was a significant predictor of mortality in our cohort, consistent with previous studies. In a study by Stefano et al.,[23] which analyzed data from 1.441.230 patients admitted to the hospital with a diagnosis of IS, AF was found to be the second strongest predictor of mortality. Patients with AF require regular anticoagulant therapy to prevent thrombus formation; however, suboptimal medication adherence for chronic conditions in our country likely contributes to

thromboembolic events and subsequent IS. Our findings on AF prevalence and its impact on post-stroke mortality are consistent with earlier reports.[24]

Patients admitted via EHS had higher mortality and complication rates compared to those presenting as outpatients. This likely reflects the poorer baseline condition of patients unable to reach the ED independently, necessitating reliance on EHS. Consequently, the higher mortality in this group is consistent with their more severe clinical status at presentation.

Regarding biochemical parameters, sodium and BUN levels were significantly higher in non-survivors. These elevations are considered indicative of dehydration.[25] Due to the retrospective design of the study, no clinical data were available to confirm dehydration at presentation, nor could differences in dehydration status between survivors and non-survivors be evaluated. However, considering the advanced age and comorbidities of many patients, dehydration due to reduced oral intake is plausible. Watso et al.[26] reported an association between dehydration and impaired cognitive function. Reduced water intake secondary to impaired consciousness or dysphagia is common in AIS patients, especially among the elderly.[27] Increased blood viscosity and hemoconcentration due to dehydration can worsen ischemic brain damage by lowering blood pressure.[28] Therefore, dehydration has been found to be associated with poor outcomes in AIS patients in numerous studies.[25,28,29]

In our study, it was observed that patients with an NIHSS score of 21 or higher had a higher frequency of mortality and complications, and they also had longer hospital stays. In a study by Fonarow et al.[30] involving 33.102 patients with recorded NIHSS scores, a strong relationship was found between increasing NIHSS score and higher 30-day mortality. Similarly, Havenon et al.[31] found that 45.9% of 6.838 patients had NIHSS scores ≥21, and mortality increased proportionally with score severity. Olum et al.[32] also demonstrated that each one-point increase in NIHSS score increased mortality risk by 6%. The higher the NIHSS score, the larger the affected cerebral tissue, and the worse the prognosis. This situation increases the risk of developing complications and prolongs the hospital stay. Therefore, irrespective of other factors, a longer hospital stay, a higher risk of complications, and extensive cerebral tissue damage will also increase the frequency of mortality.

The optimal NIHSS range for IVT is typically between 6 and 22, where the benefit-risk balance is most favorable.[33] In our study, the higher mortality rate observed in patients with an NIHSS score of 21 or higher, compared to the literature, was attributed to the administration of IVT treatment to patients with NIHSS scores above the range that requires IVT (above 22) and the influence of comorbid conditions.

Patients with a hospital stay longer than 14 days also had an increased frequency of mortality and complications. Our study examined the presence of bleeding complications (cerebral hemorrhage and gastrointestinal bleeding), which are major complications associated with thrombolytic therapy. Long-term complications of stroke (such as infections, deep vein thrombosis, or pressure ulcers) could not be assessed due to data access limitations. Studies by Vahdati et al.[21] and Heuschmann et al.[34] found a significant relationship between increased complication rates and mortality, although Vahdati et al.[21] reported no significant association between LOHS and mortality. In contrast, Kalra et al.[35] documented increased complication and mortality rates with prolonged hospitalization. These findings may reflect the delayed discharge of patients with poor prognosis, as well as the heightened risk of hospital-acquired infections over time.

We also observed a significant relationship between LOHS and complication rates in our cohort. This pattern likely reflects the extension of hospitalization in patients who develop complications. Although a direct cause-effect relationship cannot be established, our findings suggest that longer LOHS is associated with both higher complication and mortality rates.[35]

Limitations

This study has several limitations. First, it was conducted as a single-center, retrospective, cross-sectional study, which may limit the generalizability of the findings. Second, data gaps were present in the patient records maintained at our hospital. For example, blood pressure measurements at the time of ED admission were often incomplete, preventing an evaluation of blood pressure control among patients with HTN as a comorbidity.

Additionally, the small sample sizes in certain subgroups—such as patients with CHF, a history of CVD, hyperlipidemia, and those without HTN—reduced the statistical power for mortality comparisons. Furthermore, although smoking has been reported in the literature to increase the risk of AIS by up to fivefold, relevant data could not be retrieved from medical records, and therefore this risk factor could not be analyzed in our study.

CONCLUSION

This study evaluated patients presenting to the ED with AIS who received IVT treatment at a tertiary research and training hospital in Turkiye. The majority of patients were male and aged ≥65 years, with higher mortality rates observed in this age group. Although HTN was the most common comorbidity, no significant association was found between HTN and mortality. In contrast, DM and AF were associated with increased mortality. None of the comorbidities were associated with a higher frequency of complications.

The majority of patients presented to the ED via EHS, and their complication rates and mortality rates were higher compared to those who presented as outpatients. Elevated sodium and BUN levels were associated with increased mortality. Regardless of other variables, an NIHSS score of 21 and above significantly increased the risk of both complications and mortality. Hospital stays of 15 days and longer were associated with both complication and mortality rates, but independently, they significantly increased the risk of complications only. This research aims to contribute to the stroke and thrombolytic treatment data in our country and shed light on future prospective controlled randomized studies.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study protocol was approved by the Uşak University Faculty of Medicine Non-interventional Clinical Researches Ethics Committee (Date: 02.02.2023, Decision No: 65-65-19).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

Acknowledgment

The authors thank the doctors, nurses and all health professionals who contributed to the medical care of stroke patients.

REFERENCES

1. Campbell BCV, De Silva DA, Macleod MR, et al. Ischaemic stroke. *Nat Rev Dis Primers.* 2019;5(1):70. doi:10.1038/s41572-019-0118-8
2. Tsao CW, Aday AW, Almarzooq ZI, et al. Heart disease and stroke statistics-2023 update: a report from the American Heart Association. *Circulation.* 2023;147(8):e93-e621. doi:10.1161/CIR.0000000000001123
3. Feigin VL, Norrving B, Mensah GA. Global burden of stroke. *Circ Res.* 2017;120(3):439-448. doi:10.1161/CIRCRESAHA.116.308413
4. Krishnamurthi RV, Ikeda T, Feigin VL. Global, regional and country-specific burden of ischaemic stroke, intracerebral haemorrhage and subarachnoid haemorrhage: a systematic analysis of the global burden of disease study 2017. *Neuroepidemiology.* 2020;54(2):171-179. doi:10.1159/000506396
5. Chandra A, Stone CR, Du X, et al. The cerebral circulation and cerebrovascular disease III: stroke. *Brain Circ.* 2017;3(2):66-77. doi:10.4103/bc.bc_12_17
6. Auer RN, Sommer CJ. Chapter 4: histopathology of brain tissue response to stroke and injury, stroke: pathophysiology, diagnosis, and management. Elsevier Inc; 2019.
7. Jolugbo P, Ariëns RAS. Thrombus composition and efficacy of thrombolysis and thrombectomy in acute ıschemic stroke. *Stroke.* 2021;52(3):1131-1142. doi:10.1161/STROKEAHA.120.032810
8. Herpich F, Rincon F. Management of acute ischemic stroke. *Crit Care Med.* 2020;48(11):1654-1663. doi:10.1097/CCM.0000000000004597
9. Dillon GM, Stevens S, Dusenbury WL, Massaro L, Toy F, Purdon B. Choosing the correct "-ase" in acute ischemic stroke: alteplase, tenecteplase, and reteplase. *Adv Emerg Nurs J.* 2019;41(3):271-278. doi:10.1097/TME.0000000000000254
10. Warach SJ, Dula AN, Milling TJ Jr. Tenecteplase thrombolysis for acute ischemic stroke. *Stroke.* 2020;51(11):3440-3451. doi:10.1161/STROKEAHA.120.029749
11. El Husseini N, Shea KJ, Goldstein LB. Concerns for the reliability and validity of the national stroke project stroke severity scale. *Cerebrovasc Dis.* 2011;32(5):426-430.doi:10.1159/000331476
12. TR. Ministry of Health, General Directorate of Health Services. Acute ischemic stroke diagnosis and treatment guideline. Ankara; 2020.
13. Öztürk B, Özön AÖ. Ischaemic stroke and gender. *Akd Med J.* 2020;6(1):59-65. doi:10.17954/amj.2019.1579
14. Ji R, Schwamm LH, Pervez MA, Singhal AB. Ischemic stroke and transient ischemic attack in young adults: risk factors, diagnostic yield, neuroimaging, and thrombolysis. *JAMA Neurol.* 2013;70(1):51-57. doi:10.1001/jamaneurol.2013.575
15. Thrift AG, Howard G, Cadilhac DA, et al. Global stroke statistics: an update of mortality data from countries using a broad code of "cerebrovascular diseases". *Int J Stroke.* 2017;12(8):796-801. doi:10.1177/1747493017730782
16. Chantkran W, Chaisakul J, Rangsin R, Mungthin M, Sakboonyarat B. Prevalence of and factors associated with stroke in hypertensive patients in Thailand from 2014 to 2018: a nationwide cross-sectional study. *Sci Rep.* 2021;11(1):17614. doi:10.1038/s41598-021-96878-4
17. Kiyan S, Ozsarac M, Ersel M, et al. Retrospective analysis of 124 acute ischemic stroke patients who attended to the emergency department in one year period. *JAEM.* 2009;8:3. doi:10.4170/JAEM.2009.46330
18. Dolek T, Cenikli U, Bek S, Kutlu Gunergin G. Experiences of intravenous thrombolytic therapy of Muğla in acute ischemic stroke. *Med J Mugla Sitki Kocman University.* 2023;10(2):115-119. doi:10.47572/muskutd.1173741
19. Demirci Sahin A, Ustu Y, Isik D, Oztas D, Karatas Eray I, Ugurlu M. Demographic characteristics of patients with cerebrovascular disease and evaluation of preventable risk factors in primary care centers. *Ankara Med J.* 2015;15(4):196-208. doi:10.17098/amj.79435
20. Gokmen EO. Factors affecting ischemia recurrence in patients presenting to the emergency department with recurrent ischemic stroke under antiembolicor antithrombotic treatment. 2023. Access address: https://acikerisim.erbakan.edu.tr/xmlui/handle/20.500.12452/10261
21. Vahdati SS, Ala A, Hejazian S, et al. Factors affecting complications and mortality of stroke patients in stroke care unit. *Erciyes Med J.* 2022;44(2):138-142. doi:10.14744/etd.2021.71430
22. Lau LH, Lew J, Borschmann K, Thijs V, Ekinci EI. Prevalence of diabetes and its effects on stroke outcomes: a meta-analysis and literature review. *J Diabetes Investig.* 2019;10(3):780-792. doi:10.1111/jdi.12932
23. De Stefano F, Mayo T, Covarrubias C, Fiani B, Musch B. Effect of comorbidities on ischemic stroke mortality: an analysis of the National inpatient sample (NIS) database. *Surg Neurol Int.* 2021;12:268. doi:10.25259/SNI_415_2021
24. Yoshida T, Uchino S, Yokota T, Fujii T, Uezono S, Takinami M. The impact of sustained new-onset atrial fibrillation on mortality and stroke incidence in critically ill patients: a retrospective cohort study. *J Crit Care.* 2018;44:267-272. doi:10.1016/j.jcrc.2017.11.040
25. Schrock JW, Glasenapp M, Drogell K. Elevated blood urea nitrogen/creatinine ratio is associated with poor outcome in patients with ischemic stroke. *Clin Neurol Neurosurg.* 2012;114(7):881-884. doi:10.1016/j.clineuro.2012.01.031
26. Watso JC, Farquhar WB. Hydration status and cardiovascular function. *Nutrients.* 2019;11(8):1866. doi:10.3390/nu11081866
27. Crary MA, Humphrey JL, Carnaby-Mann G, Sambandam R, Miller L, Silliman S. Dysphagia, nutrition, and hydration in ischemic stroke patients at admission and discharge from acute care. *Dysphagia.* 2013;28(1):69-76. doi:10.1007/s00455-012-9414-0
28. Cortés-Vicente E, Guisado-Alonso D, Delgado-Mederos R, et al. Frequency, risk factors, and prognosis of dehydration in acute stroke. *Front Neurol.* 2019;10:305. doi:10.3389/fneur.2019.00305
29. Rowat A, Graham C, Dennis M. Dehydration in hospital-admitted stroke patients: detection, frequency, and association. *Stroke.* 2012;43(3):857-859. doi:10.1161/STROKEAHA.111.640821
30. Fonarow GC, Saver JL, Smith EE, et al. Relationship of NIHSS to 30-day mortality in medicare beneficiaries with acute ischemic stroke. *J Am Heart Assoc.* 2012;1(1):42-50. doi:10.1161/JAHA.111.000034
31. De Havenon A, Abbasi M, Yaghi S, et al. Effect of alteplase on ischemic stroke mortality is dependent on stroke severity. *Ann Neurol.* 2023;93(6):1106-1116. doi:10.1002/ana.26621
32. Olum S, Muyingo A, Wilson TL, et al. Stroke mortality outcomes in Uganda. *J Stroke Cerebrovasc Dis.* 2021;30(5):105661. doi:10.1016/j.jstrokecerebrovasdis.2021.105661

33. Ongun N. Intravenous thrombolytic therapy in acute ischemic stroke: a public hospital experience, analysis of 87 cases. *Turkish J Cerebrovasc Dis.* 2021;27(2):139-44. doi:10.5505/tbdhd.2021.2499

34. Heuschmann PU, Kolominsky-Rabas PL, Misselwitz B, et al. Predictors of in-hospital mortality and attributable risks of death after ischemic stroke: the German Stroke Registers study group. *Arch Intern Med.* 2004;164(16):1761-1768. doi:10.1001/archinte.164.16.1761

35. Kalra L, Yu G, Wilson K, Roots P. Medical complications during stroke rehabilitation. *Stroke.* 1995;26:990-994. doi:10.1161/01.str.26.6.990

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1747743

Anatolian Curr Med J. 2025;7(5):597-601

Comorbidity burden and the role of hydroxychloroquine in venous thromboembolism risk among rheumatoid arthritis patients: a retrospective case-control study

Güllü Sandal Uzun[1], Ozan Yücel[2], Yunus Kaygusuz[2], Melda Bahap Kara[3], Sedat Kiraz[4], Ali İhsan Ertenli[4], Levent Kılıç[4], Umut Kalyoncu[4]

[1]*Department of Rheumatology, Gülhane Training and Research Hospital, University of Health Sciences, Ankara, Turkiye*
[2]*Department of Internal Medicine, Faculty of Medicine, Hacettepe University, Ankara, Turkiye*
[3]*Department of Clinical Pharmacy, Faculty of Pharmacy, Hacettepe University, Ankara, Turkiye*
[4]*Division of Rheumatology, Department of Internal Medicine, Faculty of Medicine, Hacettepe University, Ankara, Turkiye*

Cite this article as: Sandal Uzun G, Yücel O, Kaygusuz Y, et al. Comorbidity burden and the role of hydroxychloroquine in venous thromboembolism risk among rheumatoid arthritis patients: a retrospective case-control study. *Anatolian Curr Med J.* 2025;7(5):597-601.

Received: 22.07.2025 ◆ **Accepted**: 17.08.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: Patients with rheumatoid arthritis (RA) are at increased risk for venous thromboembolism (VTE). This study aims to identify clinical characteristics and treatment-related factors associated with VTE in RA patients.

Methods: A retrospective analysis of 363 RA patients was conducted, including 34 RA patients with documented VTE. Demographics, comorbidities, and treatment data were compared between patients with and without VTE. Age-and sex-matched subgroup analysis and multivariate logistic regression were used to identify independent predictors of VTE.

Results: The mean age of RA patients with VTE was 66.8 (11.5) and 30 (88.2) were female. VTE was significantly associated with older age, hypertension, hyperlipidemia, and higher Charlson Comorbidity Index (CCI). In the multivariate model, higher CCI was independently associated with VTE (OR 2.84, 95% CI: 1.46–3.80, p<0.001). Hydroxychloroquine use was negatively associated with VTE (OR 0.34, 95% CI: 0.10–0.86, p=0.049). Other DMARDs and glucocorticoids did not show independent associations.

Conclusion: A high comorbidity burden significantly increases VTE risk in RA patients and HCQ use was associated with lower odds of VTE. These findings highlight the importance of individualized comorbidity management and suggest a potential role for hydroxychloroquine in thrombosis prevention in RA.

Keywords: Rheumatoid arthritis, venous thromboembolism, Charlson Comorbidity Index, hydroxychloroquine

INTRODUCTION

The incidence and prevalence of venous thromboembolism (VTE) has increased in rheumatoid arthritis (RA) due to increased chronic inflammation, older age, comorbidities, and RA-specific medications.[1-3] Epidemiological studies have found that the risk of developing VTE is 2.23 times higher than in age-gender- and comorbidity-matched control groups in patients with RA.[4] Following the approval of tofacitinib for the treatment of RA, it has been demonstrated that these treatments increase the risk of VTE and cardiovascular events.[5,6] However, real-world studies have not found a statistically significant difference in VTE incidence between patients treated with tumor necrosis factor alpha (TNF-α) inhibitors or tofacitinib.[7,8] Although studies exist on the demographic and clinical characteristics of RA patients with VTE, studies on the frequency of comorbid conditions and use of comorbidity indices are lacking. In our study, we aimed to examine the clinical and demographic characteristics of RA patients with and without VTE and to detail the factors associated with thrombosis.

METHODS

Ethics

The study has been approved by the Hacettepe University Non-interventional Clinical Researches Ethics Committee (Date: 05.10.2021, Decision No: 2021/16-31). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Patients Selection and Exclusion

This study was a retrospective single centre study. Patients with RA who were treated at the Rheumatology Outpatient

Corresponding Author: Güllü Sandal Uzun, gullusandaluzun@gmail.com

Clinics of Hacettepe University Hospitals between January to December 2021 were included in the study according to International Classification of Diseases (ICD)-10 diagnosis code. Two experienced rheumatologists reviewed the patients' medical histories, laboratory parameters, and radiographs from the hospital's electronic system, and patients with a confirmed diagnosis of RA were included in this study. Patients with RA who had a history of VTE (deep vein thrombosis (DVT) and pulmonary thromboembolism (PTE)) more than one year prior to their RA diagnosis, unconfirmed RA diagnosis, patients with incomplete medical records were excluded from the study.

Venous Thrombosis and Rheumatoid Artritis

To detect the history of VTE in RA patients; medical history, extremity venous doppler ultrasonography for the diagnosis of DVT and chest computed tomography (CT) for the diagnosis of PTE were scanned and analyzed.

Collected Data of Patients and Control Groups

We systematically reviewed the demographic and clinical characteristics, comorbid diseases/indices, laboratory parameters from the hospital's electronic medical records database. Treatments were recorded as ever and two types of drug used to treat RA: conventional synthetic disease-modifying anti-rheumatic drugs (csDMARDs), such as methotrexate, leflunomide, sulfasalazine and hydroxychloroquine, and biological disease-modifying anti-rheumatic drugs (bDMARDs), such as TNF-α and non-TNF agents, including baricitinib, tofacitinib, abatacept and rituximab were documented. We used the Charlson Comorbidity Index (CCI) to calculate the comorbidity burden.[9] We selected age-and sex-matched patients from the cohort of RA patients without VTE to serve as the "control group" (**Figure**). We compared the demographic, differences in treatment choices and clinical characteristics of RA patients with VTE with those of the control group.

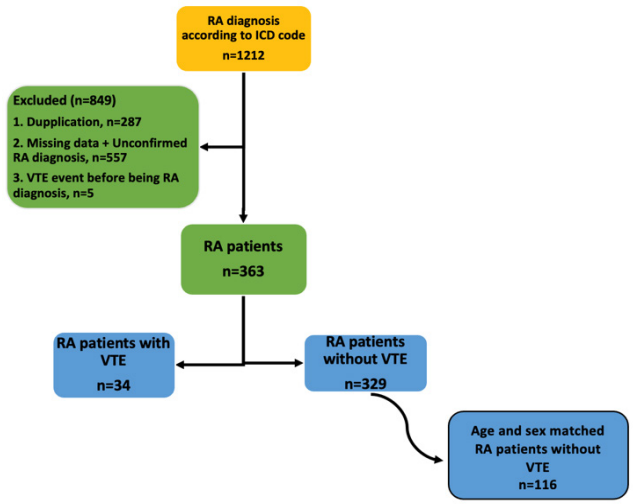


Figure. Selection of study groups

Statistical Analysis

This study utilised IBM SPSS Statistics for Windows, Version 26.0. (SPSS Inc., Chicago, IL, USA) for all statistical methods. Visual (histogram and normality plots) and analytical methods (Kolmogorov–Smirnov) were used to determine whether the data were normally distributed. For descriptive analysis, we used mean (standard deviation) and median interquartile range (IQR). We then assessed factors associated with the VTE using univariate analysis ($p<0.20$) and entered these into logistic regression by backward LR to determine independent predictors of VTE. $p<0.05$ was accepted for statistical significance.

RESULTS

Patient Characretistics

A total of 363 patients with RA confirmed diagnosis were included in the study, of whom 39 (9.4%) had a documented history of VTE. Five (12.8%) of them had VTE event before being diagnosed with RA were excluded and 34 RA patients with VTE were included in the analysis. Of the VTE patients, 16 (47%) had DVT, 16 (47%) had PTE and two (5.8%) had both. Of these patients, nine (56%) had DVT in the right lower extremity, five (8.9%) had DVT in the left lower extremity, two (1.2%) had DVT in both lower extremities, and one (6.2%) had DVT in the left upper extremity. The median time between RA diagnosis and VTE was 10.1 (1.2-50.7) years. The mean age was significantly higher in RA patients with VTE compared to those without VTE (66.8±11.5 vs. 60.6±14.9 years, $p=0.012$). Female predominance was observed in both groups, but the proportion was significantly higher in the VTE group (88.2% vs. 71.5%, $p=0.032$).

Table 1 compares RA patients with VTE to control RA patients. The prevalence of hypertension (79.4% vs. 52.7%, $p=0.005$) and hyperlipidemia (73.5% vs. 40.5%, $p=0.01$) remained significantly higher in the VTE group. The median CCI was significantly higher in patients with VTE (4.48 [1.59] vs. 3.5 [1.3], $p<0.001$), and higher proportion of patients with VTE had CCI scores ≥5 (41.2% vs. 20.7%, $p=0.016$). Rates of malignancy, coronary artery disease, and other comorbidities did not differ significantly. Glucocorticoid (96.9% vs. 62%, $p=0.02$), and methotrexate (73.5% vs. 48.2%, $p<0.001$) use was more frequent among patients with VTE. Conversely, hydroxychloroquine use was lower in the VTE group (50% vs. 80.9%, $p=0.001$). Other treatment modalities were comparable.

We examined the parameters associated with the presence of VTE by comparing them to control groups. In the univariate analysis, the following parameters had p-values less than 0.2: hypertension, hyperlipidemia, presence of malignancy, methotrexate use, hydroxychloroquine use, glucocorticoid use, and mean CCI score (**Table 2**). The factors associated with VTE in multivariate analysis are demonstrated in **Table 2**. Higher CCI score was significantly associated with VTE (OR: 2.84, 95% CI: 1.46–3.80, $p<0.001$), indicating that increased comorbidity burden strongly predicted thrombotic events. Hydroxychloroquine use was negatively associated with VTE occurrence (OR: 0.34, 95% CI: 0.10–0.86, $p=0.049$), suggesting a potential protective effect. In contrast, glucocorticoid use (OR: 0.32, $p=0.19$) and methotrexate use (OR: 0.89, $p=0.16$) were not independently associated with VTE risk after adjustment. Among cardiovascular comorbidities, both hypertension (OR: 3.6, 95% CI: 0.98–13.2, $p=0.051$) and

Table 1. Comparison of RA patients with venous thromboembolism and control RA patients

Demographic and clinical findings	Age and sex-matched RA patients without VTE n=116	RA patients with VTE n=34	p-value
Age, mean (SD)	64.4 (13.2)	66.8 (11.5)	0.28
Gender, female, n (%)	89 (76.7)	29 (85.2)	0.18
Age at the RA diagnosis, mean (SD)	49.9 (11.5)	48.3 (12.9)	0.84
Smoking history, (ever) n (%)	36 (55.4)	10/27 (37)	0.08
RF and/or ACPA positivity, n (%)	85 (78)	28 (82.4)	0.38
RF positivity, n (%)	76 (65.5)	24 (70.6)	0.37
ACPA positivity, n (%)	49 (57)	16/32 (50)	0.31
ANA positivity, n (%)	47 (57.3)	19/29 (65.5)	0.29
Comorbidities, n (%)			
Hypertension	48 (52.7)	27 (79.4)	0.005
Diabetes mellitus	21 (22.8)	6 (17.6)	0.35
Coronary artery disease	12 (12.8)	8 (23.5)	0.11
Hyperlipidemia	47 (40.5)	25 (73.5)	0.01
Thyroid diseases	18 (31.7)	12 (35.3)	0.34
Presence of malignancy	8 (6.8)	6 (17.6)	0.08
Chronic obstructive pulmonary disease	14 (15.2)	4 (12.1)	0.35
Interstitial lung disease	11 (9.4)	5 (14.7)	0.46
CCI, mean (IOR)	3.5 (1.3)	4.48 (1.59)	<0.001
CCI, subgroups, n (%) 1-2 3-4 ≥5	 34 (29.3) 58 (50) 24 (20.7)	 2 (5.9) 18 (52.9) 14 (41.2)	 0.003 0.51 0.016
Interstitial lung disease	10 (25)	5 (14.7)	0.45
Medications, n (%)			
Glucocorticoids	72 (62)	31 (96.9)	0.02
Methotrexate	56 (48.2)	25 (73.5)	<0.001
Sulfasalazine	41 (35.3)	16 (47.5)	0.07
Leflunomide	49 (42.2)	19 (55.8)	0.06
Hydroxychloroquine	93 (80.9)	17 (50)	0.001
Tumor necrosis factor inhibitors	27 (23.2)	5 (14.7)	0.27
Rituximab	14 (14.1)	7 (20.5)	0.22
Abatacept	2 (1.9)	1 (2.9)	0.14
JAK-inhibitors	5 (4.9)	1 (2.9)	0.61

RA: Rheumatoid arthritis, ANA: Antinuclear antibody, CCI: Charlson Comorbidity Index, SD: Standard deviation, IQR; Interquartile range, RF: Rheumatoid factor, ACPA: Anti-citrullinated protein antibodies, JAK: Janus kinase, VTE: Venous thromboembolism

hyperlipidemia (OR: 2.7, 95% CI: 0.88–8.6, p=0.08) showed strong trends toward increased VTE risk, though the findings were not found to be statistically significant within the multivariate model (**Table 2**).

Table 2. Predictive factors associated with VTE in RA patients

Parameters	Univariate		Multivariate	
	OR (CI)	p	OR (CI)	p
CCI, mean (SD)	3.25 (1.6-4.32)	<0.001	2.84 (1.46-3.80)	<0.001
Glucocorticoid use	0.44 (0.1-1.4)	0.06		
Methotrexate use	2.3 (0.92-7.8)	0.08		
Hydroxychloroquine use	0.26 (0.11-0.95)	0.001	0.34 (0.1-0.86)	0.049
Hyperlipidemia	1.9 (0.9-7.86)	0.09		
Hypertension	2.9 (1.4-11.6)	0.02	3.6 (0.98-13.2)	0.051

VTE: Venous thromboembolism, RA: Rheumatoid arthritis, SD: Standard deviation, CCI: Charlson Comorbidity Index, CI: Confidence interval, OR: Odds ratio

DISCUSSION

In this study, our results demonstrate that RA patients with VTE were significantly older and more likely to be female. Hypertension and hyperlipidemia were consistently more prevalent in RA patients with VTE. These findings align with existing literature indicating that conventional cardiovascular risk factors are major contributors to thrombotic events in autoimmune disease populations.

RA patients have a higher prevalence of multimorbidity compared to the age- and sex-matched general population.[10,11] Due to the lack of a standard index to scale multimorbidity in RA patients, the studies used general comorbidity indexes (Multimorbidity Weighted Index, Rheumatic Disease Comorbidity Index (RDCI), and CCI).[9,11-14] Our noteworthy finding was the association between higher comorbidity burden as measured by the CCI and the occurrence of VTE. Both univariate and multivariate analyses confirmed that increasing CCI was an independent predictor of VTE (OR: 2.84, p<0.001), highlighting the cumulative effect of multiple chronic conditions in predisposing RA patients to thrombotic complications. A recent study examining the relationship between risk factors for VTE in RA patients revealed that the mean CCI is elevated in RA patients compared to non-RA patients.[15] Özen et al.[16] also used the RDCI to determine the burden of comorbidity in their study and they highlighted that RDCI was associated with an increased risk of VTE, with an OR of 1.20 (95% CI 1.14-1.31). Consistent with our results, multimorbidity may be associated with an increased risk of VTE in RA patients. This supports the importance of considering global health status, not just RA disease activity, when evaluating VTE risk.

In our multivariate analysis, hydroxychloroquine use was inversely associated with VTE risk (OR: 0.34, p=0.049). A large real-world cohort study of RA patients showed that treatment with methotrexate led to a two-fold increased risk of VTE compared to hydroxychloroquine.[17] Studies have shown that the antithrombotic effect of hydroxychloroquine is associated with a reduced risk of thrombosis in systemic lupus erythematosus (SLE) patients.[18] A case-control study conducted in both SLE and RA patients showed that use of HCQ is associated with a decrease in the overall risk of cardiovascular events and a decrease in the development of VTE.[19] Previous studies have shown that hydroxychloroquine may have a protective effect against comorbid diseases, such as liver fibrosis. Moreover, a recently published study showed that HCQ might protect against comorbid diseases such as liver fibrosis in patients with RA.[20-22] Jung et al.[23] documented factors associated with thrombovascular events (arterial and venous thrombosis) in their study of 482 SLE patients. In

multivariate analysis, they found that older age was associated with an increased risk of VTE (OR 1.04, 95% CI 1.01–1.07), whereas antimalarial use was associated with a decreased risk of thromboembolic events. In a study of SLE patients by Akhavan et al.,[24] hydroxychloroquine use was associated with less damage in the first three years. In our study, the combined use of HCQ in RA patients was associated with a reduced risk of VTE 0.34 (0.1-0.86) (OR: 0.34, 95 CI: 0.1-0.86), consistent with the results of previous studies. While observational, this result underscores the possible thrombo-preventive benefits of hydroxychloroquine beyond its immunomodulatory role and suggests its continued use may be advantageous in high-risk RA patients. Contrary to expectations, glucocorticoid use was not independently associated with VTE risk in the multivariate model. While glucocorticoids have previously been linked to thromboembolic events due to their metabolic and vascular side effects, the widespread use of these agents across both groups in our cohort may have limited the detection of a distinct association.

Limitations

The present study has several limitations. The limited number of patients with VTE (n=34) has a detrimental effect on the statistical power of the multivariate model. Recent studies have shown a strong association between disease activity index and hospitalization duration and risk of VTE in patients with RA.[25,26] Due to retrospective design of our study, data on disease activity, functional indices, and their effects on VTE are lacking. Furthermore, there is a paucity of data regarding important parameters of VTE etiology, including prolonged hospitalization history, body-mass index, hereditary thrombophilia, and all anti-phospholipid autoantibody laboratory evaluations. The surgical history of RA patients was missing, and the temporal relationship between surgical history and VTE was unknown in our study. As a result, the relationship between these factors and VTE development could not be studied.

CONCLUSION

As a result, this study highlights that VTE in patients with RA is associated with older age, hypertension, hyperlipidaemia, and a higher burden of comorbidities. Multivariate analysis revealed that a higher CCI was independently associated with a higher risk of VTE, while the use of hydroxychloroquine was associated with a reduced risk. It is imperative that prospective studies are conducted in order to corroborate these associations and to provide a scientific basis for the development of preventive strategies for patients who are at high risk.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study has been approved by the Hacettepe University Non-interventional Clinical Researches Ethics Committee (Date: 05.10.2021, Decision No: 2021/16-31).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Omair MA, Alkhelb SA, Ezzat SE, Boudal AM, Bedaiwi MK, Almaghlouth I. Venous thromboembolism in rheumatoid arthritis: the added effect of disease activity to traditional risk factors. *Open Access Rheumatol.* 2022;14:231-242. doi:10.2147/oarrr.S284757
2. Menichelli D, Cormaci VM, Marucci S, et al. Risk of venous thromboembolism in autoimmune diseases: a comprehensive review. *Autoimmun Rev.* 2023;22(11):103447. doi:10.1016/j.autrev.2023.103447
3. Ketfi C, Boutigny A, Mohamedi N, et al. Risk of venous thromboembolism in rheumatoid arthritis. *Joint Bone Spine*. 2021;88(3):105122. doi:10.1016/j.jbspin.2020.105122
4. Lee JJ, Pope JE. A meta-analysis of the risk of venous thromboembolism in inflammatory rheumatic diseases. *Arthritis Research Therapy.* 2014;16(5):435. doi:10.1186/s13075-014-0435-y
5. Mease P, Charles-Schoeman C, Cohen S, et al. Incidence of venous and arterial thromboembolic events reported in the tofacitinib rheumatoid arthritis, psoriasis and psoriatic arthritis development programmes and from real-world data. *Ann Rheum Dis.* 2020;79(11):1400-1413. doi:10.1136/annrheumdis-2019-216761
6. Charles-Schoeman C, Fleischmann R, Mysler E, et al. Risk of venous thromboembolism with tofacitinib versus tumor necrosis factor inhibitors in cardiovascular risk-enriched rheumatoid arthritis patients. *Arthritis Rheumatol.* 2024;76(8):1218-1229. doi:10.1002/art.42846
7. Desai RJ, Pawar A, Khosrow-Khavar F, Weinblatt ME, Kim SC. Risk of venous thromboembolism associated with tofacitinib in patients with rheumatoid arthritis: a population-based cohort study. *Rheumatology (Oxford).* 2021;61(1):121-130. doi:10.1093/rheumatology/keab294
8. Duran E, Unaldi E, Bilgin E, et al. AB0416 cardiovascular event, venous thromboembolizm, and infection risk with tofacitinib in rheumatoid arthritis patients aged ≥ 60 years. *Annals of the Rheumatic Diseases*. 2022;81:1336. doi:10.1136/annrheumdis-2022-eular.4309
9. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373-383. doi:10.1016/0021-9681(87)90171-8
10. England BR, Roul P, Yang Y, et al. Burden and trajectory of multimorbidity in rheumatoid arthritis: a matched cohort study from 2006 to 2015. *Ann Rheum Dis.* 2021;80(3):286-292. doi:10.1136/annrheumdis-2020-218282
11. England BR. The multimorbidity Web in rheumatoid arthritis. *Rheumatology.* 2023;62(SI3):SI242-SI251. doi:10.1093/rheumatology/kead246
12. England BR, Sayles H, Mikuls TR, Johnson DS, Michaud K. Validation of the rheumatic disease comorbidity index. *Arthritis Care Res (Hoboken).* 2015;67(6):865-872. doi:10.1002/acr.22456

13. Wei MY, Kawachi I, Okereke OI, Mukamal KJ. Diverse cumulative impact of chronic diseases on physical health-related quality of life: implications for a measure of multimorbidity. *Am J Epidemiol*. 2016; 184(5):357-365. doi:10.1093/aje/kwv456
14. Sandal Uzun G, Tatar OD, Gezerer NE, et al. AB0278 in rheumatoid arthritis patients receiving bDMARDs, the Charlson Comorbidity Index is more pronounced than psoriatic arthritis. *Ann Rheumatic Dis*. 2022;81:1265. doi:10.1136/annrheumdis-2022-eular.3036
15. Li L, Lu N, Avina-Galindo AM, et al. The risk and trend of pulmonary embolism and deep vein thrombosis in rheumatoid arthritis: a general population-based study. *Rheumatology (Oxford)*. 2021;60(1):188-195. doi:10.1093/rheumatology/keaa262
16. Ozen G, Pedro S, Schumacher R, Simon T, Michaud K. Risk factors for venous thromboembolism and atherosclerotic cardiovascular disease: do they differ in patients with rheumatoid arthritis? *RMD Open*. 2021; 7(2):e001618. doi:10.1136/rmdopen-2021-001618
17. He M, Pawar A, Desai RJ, et al. Risk of venous thromboembolism associated with methotrexate versus hydroxychloroquine for rheumatoid arthritis: a propensity score-matched cohort study. *Semin Arthritis Rheum*. 2021;51(6):1242-1250. doi:10.1016/j.semarthrit.2021.10.001
18. Ruiz-Irastorza G, Egurbide M-V, Pijoan J-I, et al. Effect of antimalarials on thrombosis and survival in patients with systemic lupus erythematosus. *Lupus*. 2006;15(9):577-583. doi:10.1177/0961203306071872
19. Jorge A, Lu N, Choi H, Esdaile JM, Lacaille D, Avina-Zubieta JA. Hydroxychloroquine use and cardiovascular events among patients with systemic lupus erythematosus and rheumatoid arthritis. *Arthritis Care Res (Hoboken)*. 2023;75(4):743-748. doi:10.1002/acr.24850
20. Petri M. Use of hydroxychloroquine to prevent thrombosis in systemic lupus erythematosus and in antiphospholipid antibody-positive patients. *Curr Rheumatol Rep*. 2011;13(1):77-80. doi:10.1007/s11926-010-0141-y
21. Uzun GS, Bulat B, Ayan G, Kılıç L, Kalyoncu U. Liver fibrosis in inflammatory arthritis patients treated with methotrexate and hydroxychloroquine: a FIB-4 index analysis. *Int J Rheum Dis*. 2024; 27(10):e15390. doi:10.1111/1756-185x.15390
22. Meng CC, Chen DY, Chen YH, Huang WN, Chen HH. Antirheumatic drugs and the risk of nonalcoholic fatty liver disease in patients with rheumatoid arthritis: a nationwide, population-based cohort study. *Int J Rheum Dis*. 2024;27(1):e15003. doi:10.1111/1756-185x.15003
23. Jung H, Bobba R, Su J, et al. The protective effect of antimalarial drugs on thrombovascular events in systemic lupus erythematosus. *Arthritis Rheum*. 2010;62(3):863-868. doi:10.1002/art.27289
24. Akhavan PS, Su J, Lou W, Gladman DD, Urowitz MB, Fortin PR. The early protective effect of hydroxychloroquine on the risk of cumulative damage in patients with systemic lupus erythematosus. *J Rheumatol*. 2013;40(6):831-841. doi:10.3899/jrheum.120572
25. Molander V, Bower H, Frisell T, Askling J. Risk of venous thromboembolism in rheumatoid arthritis, and its association with disease activity: a nationwide cohort study from Sweden. *Ann Rheum Dis*. 2021;80(2):169-175. doi:10.1136/annrheumdis-2020-218419
26. Holmqvist ME, Neovius M, Eriksson J, et al. Risk of venous thromboembolism in patients with rheumatoid arthritis and association with disease duration and hospitalization. *Jama*. 2012;308(13):1350-1356. doi:10.1001/2012.jama.11741

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1752675

Anatolian Curr Med J. 2025;7(5):602-606

Glucose/potassium ratio as a prognostic marker in the emergency department for multiple trauma patients

Osman Lütfi Demirci, Emin Fatih Vişneci

Department of Emergency Medicine, Konya City Hospital, Konya, Turkiye

Cite this article as: Demirci OL, Vişneci EF. Glucose/potassium ratio as a prognostic marker in the emergency department for multiple trauma patients. *Anatolian Curr Med J.* 2025;7(5):602-606.

Received: 28.07.2025 ◆ **Accepted**: 18.08.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: This study aimed to evaluate the prognostic value of the glucose-to-potassium ratio (GPR) measured at emergency department (ED) admission in patients with multiple trauma who required intensive care unit (ICU) follow-up.

Methods: This retrospective study was conducted at a tertiary care center between January 1 and December 31, 2022. Adult patients admitted to the ED with multiple trauma and subsequently transferred to the ICU were included. Demographic characteristics, trauma mechanisms, laboratory values, and ICU outcomes were recorded. The predictive value of GPR for ICU mortality was analyzed using receiver operating characteristic (ROC) curve analysis. Additional comparisons were made with established biomarkers such as lactate and injury severity score (ISS).

Results: A total of 253 patients met the inclusion criteria. The most common trauma mechanisms were falls (45.7%) and traffic accidents (38.9%). Median GPR was significantly higher in non-survivors than in survivors [45.5 (30.5–63.6) vs. 31.6 (25.8–39.5), p=0.001]. ROC analysis yielded an area under the curve (AUC) of 0.712 for GPR. The optimal cut-off value was 66.9, with a sensitivity of 21% and specificity of 95%. In logistic regression analysis, GPR was identified as an independent predictor of mortality (p=0.004, Exp (B): 0.96). While lactate (AUC: 0.775) and ISS (AUC: 0.881) showed stronger predictive power, GPR remains a practical and accessible marker in the ED setting.

Conclusion: GPR is a simple, rapid, and cost-effective biomarker that may contribute to early risk stratification in multiple trauma patients. Although it should not be used in isolation for clinical decision-making, it may serve as a valuable adjunct to established prognostic tools. Further prospective and multicenter studies are warranted to validate its clinical utility.

Keywords: Glucose/potassium ratio, multiple trauma, intensive care unit, prognosis

INTRODUCTION

Trauma remains one of the leading causes of mortality worldwide, particularly affecting younger individuals.[1,2] Despite advances in medical interventions, mortality rates among patients with multiple traumatic injuries still range between 10% and 20%. Multiple trauma cases are frequently encountered in emergency departments (EDs) and require rapid and comprehensive evaluation of different organs and systems. Early diagnosis, timely treatment, and close monitoring play a critical role in reducing mortality in these patients.

Various scoring systems and imaging methods have been developed to assess prognosis in multiple trauma patients.[3,4] One of the most widely used is the injury severity score (ISS), introduced by Baker et al.,[5] which measures the overall severity of multiple injuries in the body. The ISS is calculated by summing the squares of the three highest abbreviated injury scale (AIS) scores from different body regions.[6]

An important component of the response to trauma is activation of the sympathetic nervous system, which leads to excessive release of stress-related hormones and catecholamines such as adrenaline, noradrenaline, and dopamine. Patients with multiple trauma experience this physiological stress response to a significant degree.[7]

This neurohormonal response causes marked changes in glucose and potassium metabolism. During trauma, increased catecholamine and cortisol levels stimulate hepatic gluconeogenesis and glycogenolysis, resulting in hyperglycemia. In addition, increased insulin resistance and decreased peripheral glucose utilization further contribute to this condition. Conversely, catecholamines stimulate the Na^+/K^+-ATPase pump on cell membranes, leading to intracellular potassium shift and hypokalemia.[8]

Therefore, trauma patients may present with both elevated glucose and reduced potassium levels simultaneously. The combined measure, known as the glucose-to-potassium ratio

Corresponding Author: Osman Lütfi Demirci, osman_lutfi@hotmail.com

(GPR), may reflect the severity of the stress response and the degree of metabolic imbalance. While the prognostic value of GPR has been investigated in isolated trauma settings, its significance in multiple trauma cases has not been adequately evaluated.

Accordingly, the present study aims to investigate whether GPR can serve as a predictor of mortality in multiple trauma patients admitted to the intensive care unit (ICU), in conjunction with other clinical parameters.

METHODS

Ethics

The study has received ethical approval from the Ethics Committee for Non-drug and Non-medical Device Researches at Karatay University (Date: 22.06.2023, Decision No: 2023/003). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Study Design and Participants

This retrospective study was conducted at Konya City Hospital, a tertiary care center serving approximately 50.000 patients per month, between January 1 and December 31, 2022. Data were obtained through the hospital's electronic medical record system. Clinical characteristics at ED presentation (including ISS), demographic data, ICU follow-up notes, and laboratory results were recorded.

Exclusion criteria included inaccessible medical records, incomplete clinical data, unclear medication history, absence of multiple trauma, history of diabetes, thyroid dysfunction, renal failure, severe malnutrition, extensive burns, liver cirrhosis, hemolyzed blood samples, pregnancy, age under 18 years, and the use of antihypertensive or potassium-regulating medications.

The rationale for applying these exclusion criteria was to eliminate clinical conditions that could directly alter glucose and potassium metabolism (e.g., diabetes, renal failure, thyroid dysfunction, liver cirrhosis, malnutrition, pregnancy), which might otherwise confound the study results. Similarly, medications affecting potassium balance (such as antihypertensives or potassium regulators) and conditions producing exaggerated metabolic responses (e.g., extensive burns) were excluded, since they could independently influence GPR values. This approach was intended to ensure that the findings more accurately reflected the true pathophysiological changes induced by trauma.

During the study period, a total of 724 patients admitted to the ED and subsequently transferred to the ICU were screened. Of these, 471 were excluded due to meeting one or more exclusion criteria. The most common reasons for exclusion were a history of diabetes (n=46), renal failure (n=12), incomplete clinical data (n=327), thyroid dysfunction (n=4), liver cirrhosis (n=3), pregnancy (n=6), extensive burns (n=11), and the use of antihypertensive or potassium-regulating medications (n=62). Ultimately, 253 patients were included in the study cohort.

Laboratory Analysis

Initial blood samples were collected in the ED from patients presenting with multiple trauma. Routine laboratory tests were conducted, and hematologic parameters were analyzed using an automated analyzer (Mindray BC-6000). The neutrophil-to-lymphocyte ratio (NLR) and platelet-to-lymphocyte ratio (PLR) were calculated by dividing the absolute counts of neutrophils and platelets by the absolute lymphocyte count, respectively. GPR was obtained by dividing the serum glucose concentration by the serum potassium concentration.

Statistical Analysis

Data were analyzed using IBM SPSS Statistics 23.0. Categorical variables were expressed as frequencies and percentages, while continuous variables were presented as mean±standard deviation or median (interquartile range), as appropriate. Normality was assessed using the Kolmogorov-Smirnov test.

Comparisons between groups were made using Chi-square tests for categorical variables, and the Mann–Whitney U or Student's t-test for continuous variables. ROC curve analysis was used to assess the predictive value of GPR and other parameters for ICU mortality. Univariate predictors of mortality were further evaluated using logistic regression. Parameters such as Glaskow Coma Scala (GCS) and ISS, which may introduce subjective bias, were excluded from the regression model. Model fit was assessed using the Hosmer–Lemeshow test, and the backward stepwise method was used in logistic regression analysis.

A p-value <0.05 was considered statistically significant.

RESULTS

Throughout 2022, 724 trauma patients were admitted to the ED and subsequently transferred to the ICU for further management. Of these, 253 patients met the study criteria and were included in the analysis. The demographic characteristics of these patients are summarized in **Table 1**. Regarding the mechanism of trauma, the most common cause was falls (n=113, 45.7%), followed by traffic accidents (n=96, 38.9%), occupational injuries (n=32, 13.0%), high-level falls or being pushed (n=5, 2.0%), and physical assault (n=1, 0.4%).

At admission, 189 patients (76.5%) had a GCS score of 15, 25 patients (10.5%) scored 14, and 19 patients (7.7%) scored 3. Intermediate GCS scores were less frequently observed. The ICU distribution was as follows: surgical ICU (n=147, 59.5%), anesthesiology and reanimation ICU (n=61, 24.7%), and emergency ICU (n=33, 13.4%). During follow-up, a total of 24 patients (9.7%) died in the ICU.

Among the 253 patients, 147 sustained thoracic trauma, 106 extremity trauma, 74 head injuries, 34 abdominal trauma, and 15 pelvic trauma. Seventy-five patients (30.4%) had multisystem injuries. The median ISS was 9 (range: 9–18), and the distribution of ISS values is presented in **Table 2**.

The prognostic value of GPR for predicting mortality in ICU-admitted patients was assessed using ROC curve analysis. The analysis identified a cut-off point of 66.9, corresponding to a sensitivity of 0.21 and a 1-specificity of 0.05. Median GPR

Table 1. Demographic and laboratory characteristics of the patients

Parameters*		Survival		Exitus		p	
		n	(%)	n	(%)		
Gender	Male	156	(68.1)	17	(70.8)	p= 0.05	χ2= 0.823
	Female	73	(31.9)	7	(29.2)		
Total		229	(100.0)	24	(100.0)		
			Median (IQR))		Median (IQR)		
Age		26	(36 & 77)	35	(26 & 71)	p=0.053	U=2031.5
NEU (%)		9.88	(7 & 14.78)	14.04	(9.88 & 18.81)	p=0.019	U=1896.5
LEU (%)		1.07	(0.93 & 2.26)	3.32	(1.073 & 5.67)	p=0.013	U=1846
CRP (mg/dl)		1.7	(2.1 & 30.81)	10.6	(1.6725 & 91.54)	p=0.360	U=2371.5
Lactate (mmol/L)		2.5	(1.1 & 2.7)	3.3	(2.5 & 6.7)	p<0.001	U=1206
BE (mmol/L)		-11.0	(-2.6 & 1.1)	-5.7	(-11.0 & 2.6)	p<0.001	U=780
Glucose (mg/dl)		130	(110 & 162)	185.5	(130 & 247)	p<0.001	U=1459
Potasium (mmol/L)		3.6	(3.9 & 4.5)	4.1	(3.6 & 5.1)	p=0.846	U=2611.5
GPR		30.49	(25.75 & 39.51)	45.53	(30.49 & 63.6)	p=0.001	U=1542.5
NEU/LEU		2.14	(3.3594 & 14.28)	6.33	(2.14 & 13.74)	p=0.526	U=2465
PLT/LEU		42.489	(97.446& 242.352)	66.600	(42.489 & 196.936)	p=0.007	U=1777.5
			Mean±SD		Mean±SD		
PLT			244.932±74.367		238.500±70.838	p=0.686	t=-0.404

NEU: Neutrophil, LEU: Lymphocyte, PLT: Platelet, CRP: C-reactive protein, BE: Base exes, GPR: Glucose/potassium ratio, SD: Standard deviation

Table 2. Distribution of patients according to injury severity score (ISS)

ISS	3	4	9	12	13	16	18	22	25	26	27	29	34	43	50	59
n	1	14	131	2	3	2	51	5	4	8	14	1	6	6	4	1
%	0.4	5.7	53	0.8	1.2	0.8	20.6	2	1.6	0.8	5.7	0.4	2.4	2.4	1.6	0.4

ISS: Injury severity score

was 45.5 (30.5–63.6) in non-survivors compared with 31.6 (25.8–39.5) in survivors. ROC analysis yielded an AUC of 0.712, indicating a moderate predictive accuracy for mortality (**Table 3, Figure**).

Table 3. ROC curve data for GPR and other parameters

Parameters	Area under the curve	Standard error	p	95% confidence interval
GPR	0.712	0.057	p=0.001	0.600-0.823
Lactate	0.775	0.055	p<0.001	0.666-0.883
ISS	0.881	0.054	p<0.001	0.775-0.986

ROC: Receiver operating characteristic, GPR: Glucose/potassium ratio, ISS: Injury severity score

In addition, inflammatory biomarkers were evaluated. Both the NLR and the PLR were significantly higher in non-survivors (p<0.05). Serum lactate levels were also strongly associated with mortality, with markedly elevated values in non-survivors (p<0.001). ROC curve analysis demonstrated AUC values of 0.693 for NLR, 0.642 for PLR, and 0.801 for lactate. These findings indicated that lactate was the strongest predictor of mortality; however, GPR also contributed as an independent biomarker of clinical significance.

Logistic regression analysis revealed that each one-unit increase in GPR was associated with a 0.04-fold decrease in the probability of survival (Nagelkerke R^2=0.98). Lactate and

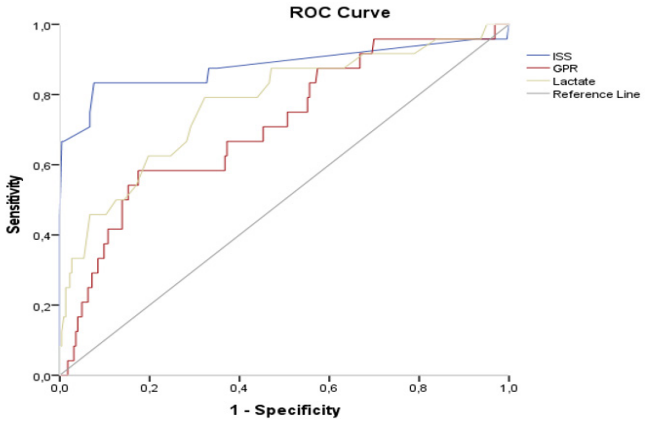


Figure. ROC curve for GPR and other parameters
ROC: Receiver operating characteristic, GPR: Glucose/potassium ratio, ISS: Injury severity score

NLR values were also found to be independently associated with mortality. Detailed results of the model are presented in **Table 4**.

Table 4. Analysis of logistic regression

	B	Sig.	Exp (B)	95% CI for Exp (B)	
				Lower	Upper
GPR	-0.37	p=0.004	0.96	0.939	0.988
BE	0.268	p<0.001	1.30	1.171	1.458

CI: Confidence interval, BE: Base exes, GPR: Glucose/potassium ratio, Hosmer–Lemeshow test: p=0.299

DISCUSSION

Multiple trauma is a severe clinical condition, typically resulting from high-energy injuries, that affects several anatomical regions and requires a multidisciplinary approach. Hemodynamic instability, metabolic stress, and inflammatory responses observed in trauma patients can lead to organ failure and increased mortality.[9-11] Therefore, there is a strong need for simple and reliable biomarkers that can predict prognosis in the early stages.

In this study, falls (45.7%) and traffic accidents (38.9%) were identified as the most common mechanisms of trauma. These findings are consistent with previous studies, such as that by Altuncu et al.,[12] which reported similar distributions. The gender distribution in our study (68.8% male and 31.2% female) also aligns with figures reported in the literature.[4,12]

Our study demonstrated that the GPR may be associated with mortality in multiple trauma patients admitted to the ED and requiring ICU care. Trauma-induced hyperglycemia and hypokalemia result in elevated GPR. These metabolic changes are primarily driven by increased sympathetic activation and catecholamine release following trauma.[7,8] Elevated glucose levels are associated with the stress response, while decreased potassium levels are linked to β-adrenergic stimulation that facilitates intracellular potassium shift.[8]

In our cohort, non-survivors exhibited markedly higher GPR values than survivors. This suggests that GPR may serve as a clinically useful prognostic biomarker in multiple trauma patients. However, our findings also indicate that while elevated GPR reflects increased mortality risk, its predictive value is limited at lower and moderate levels.

Similarly, Turan et al.[4] reported that in patients with isolated thoracoabdominal trauma, GPR was significantly associated with mortality, and its prognostic value increased when combined with the shock index. Katipoğlu et al.[13] also demonstrated a significant association between elevated GPR and mortality in trauma patients. These findings suggest that GPR may be a valid biomarker not only in isolated injuries but also in multiple trauma cases.

The systemic inflammatory response observed after trauma[14,15] may also be reflected in hematological markers such as the NLR and the PLR. However, some studies have reported that NLR tends to rise in the later stages of critical illness, limiting its utility as an early prognostic marker.[16,17] In line with this, our study found no significant association between initial NLR values and mortality. Although PLR showed some prognostic relevance, it did not demonstrate the same specificity as GPR.

On the other hand, commonly used parameters such as lactate and ISS showed stronger predictive power than GPR. Nevertheless, GPR has the advantages of being rapid, inexpensive, and widely available, making it a practical tool in clinical settings. Moreover, logistic regression analysis indicated that GPR was independently associated with mortality, supporting its potential role as a prognostic factor.

In conclusion, GPR may serve as a valuable adjunctive marker for clinicians aiming to perform rapid risk stratification in multiple trauma patients presenting to the ED. However, it should be interpreted alongside other biomarkers and clinical parameters, and not be used as a standalone tool in decision-making.

Limitations

This study has several limitations. First, it was designed as a single-center, retrospective study, which limits the generalizability of the findings. Although glucose and potassium levels were measured at the time of admission to the ED, the exact timing of trauma varied among patients and may have influenced biomarker levels.

Although we proposed that elevated GPR may be associated with increased catecholamine release, catecholamine levels were not directly measured in this study. Therefore, whether GPR truly reflects the sympathoadrenal response could not be confirmed. In addition, some clinical conditions that may affect glucose and potassium levels (such as acute stress hyperglycemia or subclinical endocrine disorders) might not have been fully excluded.

Furthermore, GPR was not directly compared with other established biomarkers, which limits the evaluation of its relative clinical value. Future studies should address these gaps using prospective, multicenter designs with larger patient populations.

CONCLUSION

This study demonstrates that GPR may be associated with mortality in multiple trauma patients. As a parameter that can be easily obtained from routine laboratory tests, it offers the advantages of being rapid, inexpensive, and widely accessible, thereby contributing to risk stratification in clinical practice. However, GPR should not be interpreted in isolation; it must be evaluated in conjunction with other biomarkers and clinical assessments.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study has received ethical approval from the Ethics Committee for Non-drug and Non-medical Device Researches at Karatay University (Date: 22.06.2023, Decision No: 2023/003).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012; 380(9859):2095-2128. doi:10.1016/S0140-6736(12)61728-0
2. Patton GC, Coffey C, Sawyer SM, et al. Global patterns of mortality in young people: a systematic analysis of population health data. *Lancet*. 2009;374(9693):881-892. doi:10.1016/S0140-6736(09)60741-8
3. de Munter L, Polinder S, Lansink KW, Cnossen MC, Steyerberg EW, de Jongh MA. Mortality prediction models in the general trauma population: a systematic review. *Injury*. 2017;48(2):221-229. doi:10.1016/j.injury.2016.12.009
4. Turan E, Şahin A. Role of glucose/potassium ratio and shock index in predicting mortality in patients with isolated thoracoabdominal blunt trauma. *Ulus Travma Acil Cerrahi Derg*. 2022;28(10):1442-1448. doi:10.14744/tjtes.2022.15245
5. Baker SP, O'Neill B, Haddon W Jr, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma*. 1974;14(3):187-196. doi:10.1097/00005373-197403000-00001
6. Keller WK, Dillihunt RC, Fenner HA, et al. Rating the severity of tissue damage: I. the abbreviated injury scale. *JAMA*. 1971;215(2):277-280. doi:10.1001/jama.1971.03180150059012
7. Offner PJ, Moore EE, Ciesla D. The adrenal response after severe trauma. *Am J Surg*. 2002;184(6):649-653. doi:10.1016/S0002-9610(02)01101-7
8. Reid JL, Whyte KF, Struthers AD. Epinephrine-induced hypokalemia: the role of beta adrenoceptors. *Am J Cardiol*. 1986;57(12):23F-27F. doi:10.1016/0002-9149(86)90884-2
9. Keel M, Trentz O. Pathophysiology of polytrauma. *Injury*. 2005;36(6):691-709. doi:10.1016/j.injury.2004.12.037
10. Butcher N, Balogh ZJ. The definition of polytrauma: the need for international consensus. *Injury*. 2009;40(Suppl 4):S12-22. doi:10.1016/j.injury.2009.10.032
11. World Health Organization. Guidelines for trauma quality improvement programmes. 2009.
12. Altunci YA, Aldemir M, Güloglu C, Üstündag M, Orak M. The effective factors in emergency department observation on hospitalization requirement and mortality in blunt trauma patients. *Eurasian J Emerg Med*. 2010;9(2):117. doi:10.4170/jaem.2010.42275
13. **13.** Katipoğlu B, Demirtas E. Assessment of serum glucose potassium ratio as a predictor for morbidity and mortality of blunt abdominal trauma. *Turk J Trauma Emerg Surg*. 2022;28(2):134. doi:10.14744/tjtes.2020.88945
14. Gary T, Pichler M, Belaj K, et al. Platelet-to-lymphocyte ratio: a novel marker for critical limb ischemia in peripheral arterial occlusive disease patients. *PLoS One*. 2013;8(7):e67688. doi:10.1371/journal.pone.0067688
15. Zahorec R. Ratio of neutrophil to lymphocyte counts—rapid and simple parameter of systemic inflammation and stress in critically ill. *Bratisl Lek Listy*. 2001;102(1):5-14.
16. Kruip MJ, Leclercq MG, van der Heul C, Prins MH, Büller HR. Diagnostic strategies for excluding pulmonary embolism in clinical outcome studies: a systematic review. *Ann Intern Med*. 2003;138(12):941-951. doi:10.7326/0003-4819-138-12-200306170-00005
17. Salciccioli JD, Marshall DC, Pimentel MA, et al. The association between the neutrophil-to-lymphocyte ratio and mortality in critical illness: an observational cohort study. *Crit Care*. 2015;19(1):13. doi:10.1186/s13054-014-0731-6

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1752081

Anatolian Curr Med J. 2025;7(5):607-612

Receptor gene expressions and their association with biochemical parameters in primary hyperparathyroidism patients

Özge Karabıyık Acar

Department of Genetics and Bioengineering, Faculty of Engineering and Natural Sciences, İstanbul Okan University, İstanbul, Turkiye

Cite this article as: Karabıyık Acar Ö. Receptor gene expressions and their association with biochemical parameters in primary hyperparathyroidism patients. *Anatolian Curr Med J.* 2025;7(5):607-612.

Received: 27.07.2025 • **Accepted**: 22.08.2025 • **Published**: 15.09.2025

ABSTRACT

Aims: Primary hyperparathyroidism (PHPT) is recognized by excessive parathyroid hormone (PTH) secretion, leading to hypercalcemia and systemic complications. This study aimed to investigate the gene expression profiles of PTH, calcium-sensing receptor (CaSR), vitamin D receptor (VDR), fibroblast growth factor receptor 1 (FGFR1), and α-Klotho from parathyroid adenomas among PHPT patients and to explore their correlations with biochemical parameters.

Methods: Parathyroid adenoma samples were obtained from 20 PHPT patients who underwent parathyroidectomy. RNA was isolated, and gene expression levels were quantified using one-step reverse transcription quantitative PCR (RT-qPCR). Fold change values were calculated utilizing gene-specific patient average ΔCt values as references. Biochemical data, including serum calcium, phosphorus, PTH, and vitamin D (25-OH), as well as gender, age, and adenoma volumes, were statistically analyzed for correlation with gene expression patterns.

Results: Gene expression analysis revealed significantly lower CaSR expression in small adenomas (<0.2 cm^3) compared to medium and large adenomas, suggesting a possible association between tumor size and early molecular deregulation. Notably, CaSR expression exhibited a positive correlation with the expression levels of both VDR and FGFR1, suggesting the presence of a coordinated regulatory network encompassing calcium and vitamin D pathways. A multiple regression model revealed that serum vitamin D and PTH gene expression serve as positive predictors of CaSR expression, whereas serum PTH levels were negatively associated. A paradoxical positive correlation was observed between serum calcium levels and PTH gene expression, which may reflect impaired feedback control due to CaSR downregulation.

Conclusion: These findings may provide new insights into the pathophysiology of PHPT, highlighting the importance of molecular profiling in understanding disease progression. The established expression patterns and their biochemical correlations may strengthen future efforts toward population-specific diagnosis and targeted therapeutic approaches, particularly in regions with high vitamin D deficiency and distinct disease severity profiles.

Keywords: Primary hyperparathyroidism (PHPT), calcium-sensing receptor (CaSR), PTH, vitamin D receptor (VDR), fibroblast growth factor receptor 1 (FGFR1), α-Klotho

INTRODUCTION

Primary hyperparathyroidism (PHPT) is a prevalent endocrine disorder characterized by the overproduction of parathyroid hormone (PTH), leading to hypercalcemia and its related systemic complications. Globally, PHPT is recognized as the third most common endocrine disease, following diabetes and thyroid-related disorders.[1] Regardless of geographical location, PHPT demonstrates an increased incidence among older populations, particularly affecting postmenopausal women (34–120 women per 100.000 person-years) when compared to men (13–36 men per 100.000 person-years).[2] While typically asymptomatic, the diagnostic criteria for classical PHPT involve serum calcium levels greater than 10.5 mg/dl and PTH levels surpassing 65 pg/ml, with the normal range being 15–65 pg/ml.[3] The pathophysiology of PHPT involves the dysregulation of calcium-regulating receptors, including the calcium-sensing receptor (CaSR), vitamin D receptor (VDR), and the α-Klotho/FGFR1 complex. According to the literature, more studies were conducted regarding the expression levels of VDR and CaSR in parathyroid adenomas.[4-7] Downregulation of CaSR and VDR is frequently observed in PHPT parathyroid adenomas and contributes to dysregulated PTH secretion.[8,9]

Expression of α-Klotho has also been identified in the parathyroid gland, and experimental evidence suggests that fibroblast growth factor 23 (FGF23) modulates PTH secretion via the FGFR1/α-Klotho axis.[10,11] Clinical investigations have reported elevated FGF23 levels in PHPT patients, with correlations to serum calcium, PTH, and

Corresponding Author: Özge Karabıyık Acar, ozgeacark@gmail.com

phosphate concentrations, although findings remain partly inconsistent.[12,13] These observations indicate that the FGFR1/α-Klotho pathway may have a distinct, yet incompletely understood role in PHPT pathophysiology, warranting further investigation at the gene expression level.

This research seeks to investigate the expression levels of PTH, CaSR, VDR, α-Klotho, and FGFR1 in parathyroid adenomas from Turkish PHPT patients. Furthermore, it aims to establish correlations between these molecular findings and biochemical parameters, including serum PTH, calcium, phosphorus, and vitamin D (25-OH) (D3) measurements, as well as adenoma size, age, and gender. Lastly, through this approach, it is intended to address the limited evidence on α-Klotho/FGFR1 signaling in PHPT. Investigating these population-specific molecular profiles may provide valuable insights for developing future treatment approaches and personalized management strategies in PHPT.

METHODS

Ethical approval was obtained from the University of Health Sciences İstanbul Training and Research Hospital Clinical Researches Ethics Committee (Date: 07.03.2025, Decision No: 52). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Patients diagnosed with PHPT who met surgical criteria and were admitted to the Endocrinology Clinic at Yeditepe University, İstanbul, Turkiye, underwent parathyroidectomy for the removal of hyperactive parathyroid tissues and were included in this study. Patients were excluded if they were younger than 18 years of age, pregnant, or had serum calcium and PTH levels within the normal reference range. Individuals with known malignancies outside the parathyroid glands were not included. In addition, patients with alternative conditions that could account for hyperparathyroidism, such as chronic renal insufficiency or sarcoidosis, were excluded. Patients with a history of previous parathyroid surgery, parathyroid carcinoma, secondary or tertiary hyperparathyroidism, or incomplete medical records were also excluded. The study included a total of 20 patients (male and female). Information details, including PTH (pg/ml), calcium (mg/dl), phosphorus (mg/dl), D3 (µg/L), gender, preoperative age, and adenoma size, were obtained from the hospital health records. Serum calcium and phosphorus concentrations were measured using photometric methods on the Roche Cobas® c 501 analyzer (Roche Diagnostics, Germany), with Roche Calcium Gen.2 and Phosphate (Inorganic) ver.2 reagent kits, respectively. Serum PTH levels were determined using an electrochemiluminescence immunoassay (ECLIA) on the Roche Cobas® e 601 analyzer (Roche Diagnostics, Germany) with the Elecsys PTH assay kit. Serum 25-hydroxyvitamin D (25-OH D3) levels were quantified using the Elecsys Vitamin D total III assay on the same platform. Adenoma size was determined by pathological assessment following surgical excision. The pathology department recorded tissue dimensions (length, width, and thickness), and tissue volumes were calculated using the ellipsoidal model with the same formula utilized in the literature: length x thickness x width x 0.52.[14,15]

RNA Isolation

All tissue samples were immediately stored at −80°C, followed by RNA isolation. Total RNA isolation was performed according to the PureLink Mini RNA Kit (Ambion-Life Technologies™, USA) instructions. Total RNA was isolated from approximately 50 mg of preserved parathyroid tissues, which were cut into small pieces and subsequently homogenized in RNA lysis buffer, including 1% ß-mercaptoethanol. Individual lysates were transferred to the spin cartridges, followed by the execution of the binding, washing, and elution steps. The RNA yield and purity were measured using a Take3 plate (BioTek instrument) with a BioTek Synergy HTX multimode reader (BioTek, US).

Reverse Transcription-Polymerase Chain Reaction

The expression levels of PTH, CaSR, α-Klotho, FGFR1, and VDR were determined using one-step RT-qPCR with SYBR Green Mastermix (2X) (Nucleogene, Turkiye). The β-actin gene was incorporated for the purposes of normalization and serving as an internal control. The QIAGEN Rotor-Gene Q-2plex device was used to perform one-step RT-qPCR for all tests in this study. Briefly, the composition of the RT-qPCR reagent mix was described in **Table 1**.

Table 1. RT-qPCR components with final concentrations

Component	Final concentration
Reaction mix (2X)	1X
Enzyme mix (40X)	1X
Primer mix (F+R)	300 nM
Nuclease-free water	
RNA template	10 ng/µL

The final volume of the RT-qPCR mixture was 16 µL, and the cycling conditions were programmed as follows: RT incubation and enzyme activation were serially performed at 50°C for 15 min and 95°C for 2 min 30 sec. Afterwards, it was cycled 45 times at 95°C for 15 sec for denaturation and 60°C for 30 sec for annealing/extension. Melt curve analysis was conducted following the instrument's recommendations. All reactions were run in triplicate. Gene expression levels were analyzed using the comparative Ct method (ΔΔCt), with the average delta Ct value of each gene across all patients serving as the reference for fold change calculations. The primer sequences are presented in **Table 2**.

Statistical Analysis

All data analyses were performed using IBM SPSS Statistics (version 27). The Shapiro–Wilk test evaluated the normality of continuous variables (n=20). For normally distributed variables, comparisons between two independent groups used the independent-samples t-test. For comparisons among three or more groups, one-way analysis of variance (ANOVA) was used. Levene's test checked the homogeneity of variances. When the equal variances assumption was met ($p \geq 0.05$),

Table 2. Target genes and the primer sequences used in the study

Target genes	5'–3' primer sequences
β-actin	F: TGATCCACATCTGCTGGAAGGT R: GACAGGATGCAGAAGGAGATTACT
PTH	F: GACATTGTATGTGAAGATGATACC R: GCTTCTTACGCAGCCATTCTA
α-Klotho	F: CCTCCTTTACCTGAAAATCAGCC R: CAGGTCGGTAAACTGAGACAGAG
CaSR	F: GGACAGCGGGAACAGGATTTGAGAG R: CCCAGTTAGTCCCGGTTCCTTCACC
VDR	F: AGCCTCAATGAGGAGCACTCCAAG R: CGGGTGAGGAGGGCTGCTGAGTA
FGFR1	F: CCAACTTAGTGAAACCCCATCT R: CCCAACAA ATACAGTCTGGTCA

PTH: Parathyroid hormone, CaSR: Calcium-sensing receptor, VDR: Vitamin D receptor, FGFR1: Fibroblast growth factor receptor 1

Tukey's honestly significant difference (HSD) post hoc test was applied; otherwise ($p<0.05$), the Games–Howell post hoc test was used to account for heterogeneity in variances. For non-normal variables, the Kruskal–Wallis H test compared group distributions.

Pearson's correlation analysis was performed to explore potential relationships between continuous variables that met the assumption of normality. For variables that were not normally distributed (e.g., PTH, D3, and α-Klotho_FC), Spearman's rank-order correlation was used instead. All statistical tests were two-tailed, and p-values less than 0.05 were regarded as statistically significant.

Categorical variables, such as gender and adenoma size classification, were analyzed using Pearson's Chi-square test of independence. In cases where more than 20% of the expected cell frequencies were less than 5, the Fisher–Freeman–Halton exact test was employed as a more appropriate alternative. Exact p-values and 99% confidence intervals were obtained using a Monte Carlo simulation with 10.000 iterations.

RESULTS

The study cohort consisted of 20 patients diagnosed with PHPT, exhibiting a mean age of 50.4±10.5 years. The biochemical analysis revealed elevated serum calcium levels (mean 11.38±0.70 mg/dl) and PTH levels (mean 132.7±59.7 pg/ml) consistent with classical PHPT. It is noteworthy that the levels of D3 were observed to be deficient, with a mean concentration of 19.8±11.5 µg/L. This finding corroborates previous reports indicating a widespread vitamin D insufficiency within the Turkish population.[16] The mean phosphorus level was recorded as 2.68±0.49 mg/dl. The mean adenoma volume was 0.51 cm^3 (range: 0.12–1.95 cm^3) (**Table 3**). Gender-based comparisons (**Table 4**) utilizing an independent-samples t-test indicated that male patients exhibited significantly higher calcium levels (11.95±0.84 mg/dl) compared to females (11.20±0.57 mg/dl; $t(18)=2.313$, $p=0.032$). The effect size was moderate (Cohen's $d=0.64$), signifying a practically meaningful difference. While no significant differences were identified in phosphorus levels or the expression of key genes (PTH_FC, FGFR1_FC, VDR_FC, CaSR_FC) across genders ($p>0.05$), there were small effect sizes observed for PTH_FC ($d=0.37$) and CaSR_FC ($d=0.32$). These findings are considered preliminary and should be interpreted with caution. To confirm their potential biological importance, further studies with larger sample sizes and greater statistical power are needed. For the variables that violated the normality assumption (PTH, D3, and α-Klotho_FC), the Mann–Whitney U test revealed no significant gender-based differences: PTH ($U=33.000$, $p=0.563$), D3 ($U=34.500$, $p=0.649$), and α-Klotho_FC ($U=34.000$, $p=0.620$).

Table 3. Descriptive clinical and biochemical parameters of the study cohort (n=20). Values presented include the mean, median, minimum, and maximum for pre-operative age, serum PTH (pg/ml), serum calcium (mg/dl), serum phosphorus (mg/dl), serum D3 (µg/L), and adenoma volume (cm^3)

	Age	PTH (pg/ml)	Calcium (mg/dl)	Phosphorus (mg/dl)	D3 (µg/L)	Adenoma volume (cm^3)
Mean	50	132.72	11.38	2.69	19.79	0.51
Median	50	118.00	11.29	2.70	18.00	0.27
Minimum	26	65.30	10.22	1.93	3.00	0.12
Maximum	69	335.30	13.10	3.53	58.00	1.95

PTH: Parathyroid hormone

The analysis of gene expression patterns, utilizing a one-way ANOVA, revealed significant variation in CaSR_FC levels across adenoma size categories ($F(2, 18)=3.664$, $p=0.046$) (**Table 5**). The post-hoc analysis using the Games–Howell test demonstrated that small adenoma groups (<0.2 cm^3) displayed significantly lower CaSR_FC expression compared to medium (0.2–0.5 cm^3; mean difference=–3.81909, $p=0.048$) and large (>0.5 cm^3; mean difference=–4.69167, $p=0.014$) adenoma groups, suggesting a potential relationship between adenoma size and CaSR dysregulation. Nevertheless, no statistically significant difference was observed between the medium and large adenoma groups ($p=0.756$).

Correlation analysis between biochemical and molecular variables was shown in **Table 6**. Pearson's correlation analysis identified several significant associations, including a positive relationship between calcium levels and PTH_FC ($r=0.520$, $p=0.016$), suggesting that increased calcium levels are correlated with higher PTH gene expression. Additionally, FGFR1_FC was strongly correlated with VDR_FC ($r=0.616$, $p=0.003$), and CaSR_FC showed positive correlations with both FGFR1_FC ($r=0.434$, $p=0.049$) and VDR_FC ($r=0.605$, $p=0.004$), implying potential regulatory or functional relationships. No other significant Pearson correlations were identified, and age did not exhibit a significant correlation with any continuous variables (all $p>0.05$). Spearman's rank-order correlations among non-normally distributed variables revealed no significant associations. Specifically, PTH was not significantly correlated with D3 ($\rho=-0.064$, $p=0.783$) or α-Klotho_FC ($\rho=-0.134$, $p=0.563$), and no association was observed between D3 and α-Klotho_FC ($\rho=-0.071$, $p=0.759$).

A multiple linear regression was performed to identify predictors of CaSR_FC expression. Predictor variables included adenoma volume, age, calcium, phosphorus, PTH, D3, PTH_FC, VDR_FC, FGFR1_FC, and α-Klotho_FC. The model was statistically significant, $F(10, 10)=4.288$, $p=0.015$, explaining 81.1% of the variance in CaSR_FC expression ($R^2=0.811$; adjusted $R^2=0.622$). Model diagnostics confirmed that the residuals were approximately normally

Table 4. Gender-based comparisons of biochemical and molecular parameters

Variable	Male mean±SD	Female mean±SD	Test statistic (df)	p-value	Effect size (Cohen's d)
Calcium (mg/dl)	11.96±0.84	11.20±0.57	t (19)=2.313	0.032*	0.64 (moderate)
Phosphorus (mg/dl)	2.69±0.64	2.68±0.45	t (19)=0.289	0.776	0.09 (trivial)
PTH (pg/ml)	161.52±102.65	123.73±39.81	U=33.000	0.563	0.49 (medium)
D3 (µg/L)	18.72±8.43	10.13±12.57	U=34.500	0.649	0.80 (large)
PTH_FC	2.38±2.45	−0.74±12.57	t (18)=1.392	0.179	0.37 (small)
CaSR_FC	0.81±3.84	−0.25±3.03	t (18)=0.785	0.445	0.32 (small)
α-Klotho_FC	−0.12±4.22	0.04±2.72	U=34.000	0.620	0.05 (trivial)

Independent-samples t-test for normally distributed variables; Mann–Whitney U test for non-normally distributed variables. * p<0.05 was considered statistically significant. SD: Standard deviation, PTH: Parathyroid hormone, CaSR: Calcium-sensing receptor

Table 5. Comparison of CaSR_FC expression across adenoma size categories

Comparison	Mean difference	p-value
Small vs. medium	−3.81909	0.048*
Small vs. large	−4.69167	0.014*
Medium vs. large	0.87258	0.756

CaSR: Calcium-sensing receptor, Adenoma size: small (<0.2 cm^3), medium (0.2–0.5 cm^3), large (>0.5 cm^3), Post-hoc comparisons were conducted using the Games–Howell test. * p<0.05 was considered statistically significant

Table 6. Correlation analysis between biochemical and molecular variables

Variable pair	Correlation coefficient	p-value
Calcium vs. PTH_FC	r=0.520	0.016*
FGFR1_FC vs. VDR_FC	r=0.616	0.003*
CaSR_FC vs. FGFR1_FC	r=0.434	0.049*
CaSR_FC vs. VDR_FC	r=0.605	0.004*
PTH vs. D3	ρ=−0.064	0.783
PTH vs. α-Klotho_FC	ρ=−0.134	0.563
D3 vs. α-Klotho_FC	ρ=−0.071	0.759

PTH: Parathyroid hormone, CaSR: Calcium-sensing receptor, VDR: Vitamin D receptor, FGFR1: Fibroblast growth factor receptor 1. Pearson's correlation-for normally distributed variables; Spearman's correlation-for non-normally distributed variables. * p<0.05 was considered statistically significant.

distributed, homoscedastic, and that no multicollinearity issues were present (VIFs<2.55). Three variables emerged as significant predictors: PTH levels showed a negative association (B=−0.029, β=−0.546, p=0.017), while D3 (B=0.114, β=0.414, p=0.034) and PTH_FC (B=0.585, β=0.715, p=0.007) demonstrated positive relationships with CaSR_FC expression. The remaining variables, including adenoma volume, age, calcium, phosphorus, VDR_FC, FGFR1_FC, and α-Klotho_FC, did not reach statistical significance but contributed to the overall model. To assess model adequacy, a post-hoc power analysis was conducted using G*power (version 3.1.9.7). Based on the observed effect size (f^2=4.29), 10 predictors, and a sample size of 20, the achieved power was 0.997, indicating a very high probability of detecting a true effect.

DISCUSSION

This study provides new perspectives on the biochemical and molecular characteristics of PHPT in a Turkish patient population. Our cohort demonstrated elevated mean PTH and calcium levels, consistent with previous reports from Turkiye. For instance, a 20-year single-center study reported a mean serum calcium of 11.9±2.2 mg/dl and a mean PTH of 467±78 pg/ml among 190 PHPT patients, of whom 72% were asymptomatic at diagnosis.[17] Similarly, a comparative study between Turkiye (Bursa) and the United States cohorts demonstrated significantly higher PTH concentrations (546±75 pg/ml vs. 146±75 pg/ml) and larger adenoma sizes (25.2±1.18 mm vs. 17.5±1.18 mm) among Turkish patients.[18] Furthermore, vitamin D deficiency has been demonstrated to be notably common in Turkish nationals (75%; <50 nmol/l), contributing to more severe biochemical profiles and secondary HPT.[19] These observations support our findings and highlight that Turkish PHPT patients may present with more pronounced biochemical alterations compared to Western populations.

The findings showed complex regulation of PTH, CaSR, VDR, FGFR1, and α-Klotho gene expressions in PHPT adenomas, consistent with previous literature. In this study, CaSR expression (CaSR_FC) was determined to be significantly downregulated in small adenomas (<0.2 cm^3) in comparison to medium and large adenomas, suggesting a potential correlation between the size of adenomas and CaSR dysregulation. This finding is consistent with previous research identifying CaSR downregulation as a hallmark of PHPT. Yano et al.[4] demonstrated that a reduction in CaSR and VDR expression was associated with higher proliferation markers and increased adenoma mass, indicating that CaSR may play an important role in the dynamics of tumor growth. Yet, other studies, such as Varshney et al.,[20] did not establish a direct correlation with tumor weight; instead, they confirmed reduced CaSR mRNA levels in adenomas among the Asian Indian populations. The observation that the lowest CaSR expression was detected in the smallest tumors may tentatively suggest that the loss of CaSR could represent an early alteration in the progression of tumor development, occurring before mass expansion. However, given the cross-sectional nature of this study and the limited number of samples, this interpretation should be considered exploratory and requires validation in larger, longitudinal studies. This observation is also corroborated by immunohistochemical verification from an autopsy-based study, which showed that reduced membrane CaSR expression occurs not only in adenomas but also in the normal parathyroid rim tissue, implying an early and diffuse CaSR suppression in PHPT pathogenesis.[21]

Notably, CaSR_FC exhibited a positive correlation with both VDR_FC and FGFR1_FC. Our results demonstrating downregulation of CaSR and its significant associations with VDR and FGFR1 expression are in line with the immunohistochemical findings of Latus et al.,[22] who reported markedly reduced protein expression of CaSR and VDR in parathyroid glands of PHPT patients compared with controls, while FGFR expression remained unchanged. Interestingly, while Latus et al.[22] did not observe significant correlations of Klotho with serum calcium levels, our dataset revealed interrelationships between CaSR_FC, VDR_FC, and FGFR1_FC at the transcriptional level, suggesting that the molecular cross-talk between these pathways may manifest differently at the mRNA versus protein expression level. This may suggest the presence of a coordinated regulatory network involving calcium sensing, vitamin D signaling, and the FGF23/FGFR1/Klotho pathways in parathyroid adenomas. While the existing literature regarding FGFR1 and α-Klotho expression in PHPT is limited, the data presented indicate a possible interconnected cross-talk across these pathways, aligning with studies that highlight a reduction in CaSR, VDR, and FGFR1-Klotho receptors in chronic hyperplasia.[23] In another SHPT study, decreased α-Klotho and FGFR1 expression in hyperplastic parathyroids were related, inversely, to parathyroid volume,[11] supporting the observation of lower CaSR_FC in smaller adenomas and its interplay with FGFR1_FC. Additionally, low vitamin D status has been associated with increased adenoma weight and serum PTH levels in Turkish patients,[24] Similarly, our cohort demonstrated biochemical severity, featuring elevated levels of calcium and PTH, which corresponds with the national characteristics. The finding of a positive association between D3 and CaSR_FC is consistent with the literature, suggesting that active vitamin D promotes CaSR transcription through vitamin D response elements in the CaSR promoter.[25]

Interestingly, the significant positive correlation between serum calcium levels and PTH gene expression may appear paradoxical under normal physiological regulation, though it aligns well with the pathophysiological mechanisms underlying PHPT. Under physiological conditions, raised levels of extracellular calcium inhibit PTH transcription and secretion by activating the CaSR in parathyroid cells. However, in PHPT, CaSR expression is typically downregulated or functionally impaired. This disruption of the negative feedback mechanism permits the ongoing PTH synthesis, despite hypercalcemia.[4] The excessive production of PTH contributes to further bone resorption via osteoclast activation, thereby sustaining a cycle of hypercalcemia and hormonal dysregulation.[26] Therefore, the identified positive correlation possibly indicates the autonomous nature of adenomatous parathyroid tissue, which continues to express PTH mRNA despite elevated serum calcium levels. Lastly, the significantly higher serum calcium levels observed in male patients may represent a preliminary observation of potential gender-related variation in PHPT. Given the small cohort size, this finding should be interpreted with caution and considered exploratory until validated in larger studies.

Limitations

This study presents valuable insights into gene expression dynamics in PHPT patients; however, several limitations must be acknowledged. In light of these limitations, the interpretations presented herein should be regarded as exploratory and hypothesis-generating rather than definitive.

First, the sample size was limited to 20 parathyroid adenoma tissues, which may restrict the generalizability of the findings. Secondly, due to ethical restrictions, normal parathyroid tissues from healthy individuals could not be included as direct controls. While such samples are difficult to obtain, it is also important to recognize that even histologically "normal" parathyroid tissue adjacent to adenomas in PHPT patients or thyroid-related disease patients may not reflect truly unaffected physiology. Therefore, the use of such adjacent tissue as a control could yield misleading results. Instead, the average ΔCt values derived from the entire patient cohort for each gene were used as reference baselines for fold-change calculations. While this method allowed for internal normalization, it may obscure subtle expression variations that could be better highlighted through comparison with truly unaffected tissue. In addition, clinical parameters such as serum creatinine levels, detailed comorbidity data, and medication histories were not consistently available for all patients and were therefore not included. While none of the patients had known chronic renal insufficiency or major comorbidities, the absence of these parameters should be recognized as a limitation.

Lastly, this investigation represents the initial molecular profiling phase of a broader research project titled "Investigating efficacy of computationally repurposed drugs Cucurbitacin I, DG 041, IMD 0354 in in vitro PHPT Model regarding gene and protein pathways". This project is also a continuation of a previously published study.[27] Future studies will build upon this dataset by testing the effects of candidate drugs in functional in vitro PHPT models.

CONCLUSION

This study provides the first comprehensive qPCR-based profiling of PTH, CaSR, VDR, FGFR1, and α-Klotho expression, with a specific focus on PHPT patients from Turkiye. It contributes to the growing body of literature that represents distinct clinical and molecular features of PHPT. The findings highlight the downregulation of CaSR expression in smaller adenomas, suggesting that CaSR suppression may be an early event in parathyroid tumors. The observed correlations among CaSR, VDR, and FGFR1 gene expressions further point to a potentially coordinated regulatory network involving calcium, vitamin D, and FGF23/Klotho signaling pathways in parathyroid adenomas. Additionally, the paradoxical positive correlation between serum calcium and PTH gene expression supports the notion of impaired feedback inhibition attributed to CaSR dysfunction, a hallmark of PHPT pathophysiology. The results presented, along with the spotted gender-related differences in calcium levels and the observed molecular profiles, emphasize the

importance of incorporating both biochemical and gene expression parameters in the investigation of PHPT. Further research involving larger and more diverse populations is warranted to validate these findings and to explore their potential clinical implications, including early diagnosis and personalized treatment strategies.

ETHICAL DECLARATIONS

Ethics Committee Approval

Ethical approval was obtained from the University of Health Sciences İstanbul Training and Research Hospital Clinical Researches Ethics Committee (Date: 07.03.2025, Decision No: 52).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

This work was supported by İstanbul Okan University Scientific Research Projects Coordination under the Grant Number OBAP2024010002.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

Acknowledgement

The author gratefully acknowledges Prof. Dr. Erhan Ayşan and Assistant Prof. Dr. Servet Tunoğlu for their invaluable support.

REFERENCES

1. Kim SJ, Shoback DM. Sporadic primary hyperparathyroidism. *Endocrinol Metab Clin North Am*. 2021;50(4):609-628. doi:10.1016/j.ecl.2021.07.006
2. Madkhali T, Alhefdhi A, Chen H, Elfenbein D. Primary hyperparathyroidism. *Ulus Cerrahi Derg*. 2016;32(1):58-66. doi:10.5152/UCD.2015.3032
3. Fedorova T, Hagglund K, Hawasli A. The normohormonal primary hyperparathyroidism; a surgical dilemma pre and intra-operatively. *Am J Surg*. 2022;223(3):589-591. doi:10.1016/j.amjsurg.2022.01.002
4. Yano S, Sugimoto T, Tsukamoto T, et al. Decrease in vitamin D receptor and calcium-sensing receptor in highly proliferative parathyroid adenomas. *Eur J Endocrinol*. 2003;148(4):403-411. doi:10.1530/eje.0.1480403
5. Li X, Lu Y, Zhang L, et al. Primary and secondary hyperparathyroidism present different expressions of calcium-sensing receptor. *BMC Surg*. 2023;23(1):31. doi:10.1186/s12893-023-01928-5
6. Rao S, Han, Phillips ER, Palnitkar S, Parfitt. Reduced vitamin D receptor expression in parathyroid adenomas: implications for pathogenesis. *Clin Endocrinol*. 2000;53(3):373-381. doi:10.1046/j.1365-2265.2000.01081.x
7. Singh P, Bhadada SK, Dahiya D, et al. Reduced calcium sensing receptor (CaSR) expression is epigenetically deregulated in parathyroid adenomas. *J Clin Endocrinol Metab*. 2020;105(9):3015-3024. doi:10.1210/clinem/dgaa419
8. Hendy GN, Canaff L. Calcium-sensing receptor gene: regulation of expression. *Front Physiol*. 2016;7:394. doi:10.3389/fphys.2016.00394
9. Carling T, Rastad J, Szabó E, Westin G, Åkerström Gr. Reduced parathyroid vitamin D receptor messenger ribonucleic acid levels in primary and secondary hyperparathyroidism*. *J Clin Endocrinol Metab*. 2000;85(5):2000-2003. doi:10.1210/jcem.85.5.6607
10. Kawakami K, Takeshita A, Furushima K, et al. Persistent fibroblast growth factor 23 signalling in the parathyroid glands for secondary hyperparathyroidism in mice with chronic kidney disease. *Sci Rep*. 2017;7:40534. doi:10.1038/srep40534
11. Yan J, Jingbo C, Wang D, et al. A correlation between decreased parathyroid α-Klotho and fibroblast growth factor receptor 1 expression with pathological category and parathyroid gland volume in dialysis patients. *Int Urol Nephrol*. 2015;47(4):701-706. doi:10.1007/s11255-015-0917-0
12. Yamashita H, Yamashita T, Miyamoto M, et al. Fibroblast growth factor (FGF)-23 in patients with primary hyperparathyroidism. *Eur J Endocrinol*. 2004;151(1):55-60. doi:10.1530/eje.0.1510055
13. Kobayashi K, Imanishi Y, Miyauchi A, et al. Regulation of plasma fibroblast growth factor 23 by calcium in primary hyperparathyroidism. *Eur J Endocrinol*. 2006;154(1):93-99. doi:10.1530/eje.1.02053
14. Özçelik S, Baş S, Çelik M, et al. Relationship of parathyroid adenoma volume with preoperative biochemical parameters. *J Surg Med*. 2019;3(1):1-3. doi:10.28982/josam.435391
15. Korkmaz H. The diagnostic value of parathormone washout in Tc-99m MIBI negative primary hyperparathyroidism cases. Tc-99m MIBI negatif primer hiperparatiroidizm olgularında parathormon yıkamanın tanısal değeri. *Acta Medica Alanya*. 2021;5(1):61-65. doi:10.30565/medalanya.853038
16. Alpdemir M, Alpdemir MF. Meta analysis vitamin D deficiency status in Turkey: a meta-analysis. *Int J Med Biochem*. 2019;2(3):118-131.
17. Usta A, Alhan E, Cinel A, Türkyılmaz S, Erem C. A 20-year study on 190 patients with primary hyperparathyroidism in a developing country: Turkey experience. *Int Surg*. 2015;100(4):648-655. doi:10.9738/intsurg-d-14-00094.1
18. Kirdak T, Duh Q-Y, Kebebew E, Clark OH. Do patients undergoing parathyroidectomy for primary hyperparathyroidism in San Francisco, CA, and Bursa, Turkey, differ? *Am J Surg*. 2009;198(2):188-192. doi:10.1016/j.amjsurg.2008.09.031
19. Erkal MZ, Wilde J, Bilgin Y, et al. High prevalence of vitamin D deficiency, secondary hyperparathyroidism and generalized bone pain in Turkish immigrants in Germany: identification of risk factors. *Osteoporos Int*. 2006;17(8):1133-1140. doi:10.1007/s00198-006-0069-2
20. Varshney S, Bhadada SK, Saikia UN, et al. Simultaneous expression analysis of vitamin D receptor, calcium-sensing receptor, cyclin D1, and PTH in symptomatic primary hyperparathyroidism in Asian Indians. *Eur J Endocrinol*. 2013;169(1):109-116. doi:10.1530/eje-13-0085
21. Agarwal S, Kardam S, Chatterjee P, et al. CaSR expression in normal parathyroid and PHPT: new insights into pathogenesis from an autopsy-based study. *J Endocrinol Invest*. 2022;45(2):337-346. doi:10.1007/s40618-021-01646-w
22. Latus J, Lehmann R, Roesel M, et al. Involvement of α-Klotho, fibroblast growth factor-, vitamin-D- and calcium-sensing receptor in 53 patients with primary hyperparathyroidism. *Endocrine*. 2013;44(1):255-263. doi:10.1007/s12020-013-9881-6
23. Naveh-Many T, Volovelsky O. Parathyroid cell proliferation in secondary hyperparathyroidism of chronic kidney disease. *Int J Mol Sci*. 2020;21(12):4332. doi:10.3390/ijms21124332
24. Özbey N, Erbil Y, Ademoğlu E, Özarmağan S, Barbaros U, Bozbora A. Correlations between Vitamin D status and biochemical/clinical and pathological parameters in primary hyperparathyroidism. *World J Surg*. 2006;30(3):1658. doi:10.1007/s00268-005-0239-y
25. Canaff L, Hendy GN. Human calcium-sensing receptor gene. Vitamin D response elements in promoters P1 and P2 confer transcriptional responsiveness to 1,25-dihydroxyvitamin D. *J Biol Chem*. 2002;277(33):30337-30350. doi:10.1074/jbc.M201804200
26. Levine BS, Rodríguez M, Felsenfeld AJ. Serum calcium and bone: effect of PTH, phosphate, vitamin D and uremia. *Nefrología (English Edition)*. 2014;34(5):658-669. doi:10.3265/Nefrologia.pre2014.Jun.12379
27. Kubat Öktem E, Yazar M, Aysan E, Karabıyık Acar Ö. Computational drug repurposing for primary hyperparathyroidism. *Mol Cell Endocrinol*. 2024;583:112159. doi:10.1016/j.mce.2024.112159

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1739247

Anatolian Curr Med J. 2025;7(5):613-617

The relationship between sleep quality and inflammatory and biochemical parameters in hemodialysis patients

Arzu Akgül[1], Neriman Sıla Koç[1], Berrak Itır Aylı[1], Hakkı Öztürk[2], Mehmet Deniz Aylı[1]

[1]*Department of Nephrology, Ankara Etlik City Hospital, Ankara, Turkiye*

[2]*Afyonkarahisar Provincial Health Services Directorate, Provincial Health Director, Afyonkarahisar, Turkiye*

Cite this article as: Akgül A, Koç NS, Aylı BI, Öztürk H, Aylı MD. The relationship between sleep quality and inflammatory and biochemical parameters in hemodialysis patients. *Anatolian Curr Med J.* 2025;7(5):613-617.

Received: 10.07.2025 ◆ **Accepted**: 22.08.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: This study aimed to evaluate the relationship between sleep quality and inflammatory and biochemical parameters in hemodialysis patients.

Methods: In this cross-sectional, single-center study conducted in January 2025, a total of 195 patients aged 18 and older who had been undergoing hemodialysis treatment for at least six months were included. Sleep quality was assessed using the Pittsburgh Sleep Quality Index (PSQI); scores ≤5 were classified as "good sleep," and >5 as "poor sleep." Demographic, clinical, and laboratory data were retrospectively collected. Following univariate comparisons, logistic regression analysis was performed to determine independent predictors of poor sleep quality.

Results: Poor sleep quality was observed in 55% of the patients. In the univariate analysis, CRP, leukocytes, and CRP/albumin ratio were higher in poor sleepers, while hemoglobin, albumin, and HALP scores were lower ($p<0.05$). Multivariate logistic regression analysis revealed that elevated CRP (OR 1.032; 95% CI 1.001–1.064; $p=0.043$) and low hemoglobin (OR 0.481; 95% CI 0.350–0.660; $p<0.001$) were independent predictors of poor sleep quality.

Conclusion: Sleep disorders in hemodialysis patients are closely associated with systemic inflammation and anemia. CRP and hemoglobin levels are strong independent determinants of poor sleep quality. It is recommended that these parameters be routinely monitored in clinical practice and that interventions targeting inflammation and anemia be developed.

Keywords: Hemodialysis, sleep quality, CRP, hemoglobin, biomarker

INTRODUCTION

End-stage renal disease (ESRD) is a multidisciplinary health issue that significantly affects patients' physical, mental and social health as well as their quality of life. According to data from the Turkish Society of Nephrology at the end of 2023, 70.88% of patients receiving renal replacement therapy (RRT) in Turkiye are treated with hemodialysis—the most common treatment modality.[1] The chronic nature of hemodialysis and its accompanying complications significantly impact patients' physical, psychological, and social functioning, one of the major manifestations of which is sleep disturbances.[2,3]

Sleep disorders are frequently observed but often overlooked clinical conditions in hemodialysis patients.[4] Studies have reported various sleep problems—such as difficulty falling asleep, frequent awakenings, non-restorative sleep, excessive daytime sleepiness, and restless leg syndrome—in up to 50–85% of hemodialysis patients.[2,4-6] Poor sleep quality has been associated with increased cardiovascular events, cognitive decline, and reduced overall quality of life.[7] Therefore, early identification and management of sleep disturbances in hemodialysis patients should be prioritized.

The etiology of sleep disorders in hemodialysis patients is multifactorial. Underlying pathophysiological mechanisms include circadian rhythm disruptions, sympathetic-vagal imbalance, reduced melatonin secretion, vitamin D deficiency, altered cerebral hemodynamics, and dialysis-induced vascular stress. These disruptions are also associated with frontal lobe dysfunction and significant impairments in executive functions.[5,8]

In addition to these physiological changes, several studies have shown significant associations between sleep disorders and biochemical/inflammatory parameters such as low hemoglobin, hypoalbuminemia, elevated C-reactive protein (CRP), hyperphosphatemia, and leukocytosis.[3,9]

This study aims to evaluate the relationship between sleep quality and demographic, clinical, biochemical, and inflammatory parameters—including hemoglobin albumin lymphocyte platelet (HALP) score, CRP, and albumin—in ESRD patients who have been on hemodialysis for at least six months. It is one of the few studies to investigate the

Corresponding Author: Arzu Akgül, arzuakgul@gmail.com

association between sleep quality and composite biomarkers like HALP score in the Turkish hemodialysis population and thus makes a valuable contribution to the literature.

METHODS

The study has been approved by the Scientific Researches Evaluation and Ethics Committee of Ankara Etlik City Hospital (Date: 18.12.2024, Decision No: 2024-1230) and conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants.

This cross-sectional study was conducted among ESRD patients who had been receiving hemodialysis treatment for at least six months and attended the Adult Nephrology Outpatient Clinic at Ankara Etlik City Hospital. The study was conducted in January 2025. A total of 195 individuals aged 18 years or older, cognitively capable of completing the questionnaire, and without active infection, malignancy, or altered consciousness were included.

To assess sleep patterns over the past month, the Pittsburgh Sleep Quality Index (PSQI) was administered via face-to-face interviews. The PSQI was developed by Buysse et al.[10] in 1989, and adapted into Turkish by Ağargün et al.[11] in 1996, with a Cronbach's alpha coefficient of 0.8. The original scale includes 24 questions—19 self-reported and 5 completed by a bed partner. In this study, only the 19 self-reported items were used. These evaluate sleep across seven components: subjective sleep quality, sleep latency, sleep duration, habitual sleep efficiency, use of sleep medication, sleep disturbances, and daytime dysfunction. The scale is scored out of 21, with higher scores indicating poorer sleep quality. A total score of 5 or below is considered good sleep quality, whereas scores above 5 indicate poor sleep quality.[11]

Additionally, a data collection form was administered to gather participants' demographic data (age, sex) and clinical characteristics (duration of dialysis, comorbidities). Current biochemical and hematological data were retrospectively retrieved from the hospital's electronic medical records. Patients without available laboratory data from the past month were excluded from the study. Inflammatory markers including CRP, albumin, leukocytes, neutrophils, and lymphocytes were recorded. HALP score and CRP/albumin ratio were calculated.

The HALP score was calculated using the following formula:

HALP=(hemoglobin [g/L]×albumin [g/L]×(lymphocyte [10^9/L]/platelet [10^9/L])

Statistical Analysis

All data were analyzed using SPSS (Statistical Package for the Social Sciences) version 23.0. Descriptive statistics for numerical variables were presented as mean±standard deviation (SD), and categorical variables were presented as numbers and percentages (%). For comparisons between groups, independent samples t-test was used for normally distributed variables and the Mann-Whitney U test for non-normally distributed variables. Chi-square or Fisher's exact test was applied to analyze categorical data.

Spearman correlation analysis was conducted to assess the relationship between sleep quality and laboratory/clinical parameters. To identify independent risk factors affecting sleep quality, Both univariate and multivariable logistic regression analyses were performed. the results are presented as odds ratios (OR) with corresponding 95% confidence intervals (CI) and p-values. multivariable linear regression models were constructed. A p-value <0.05 was considered statistically significant.

RESULTS

A total of 195 hemodialysis patients were included in the study. The mean age of participants was 59.1±13.5 years, with 38.9% being female. The average duration of chronic kidney disease (CKD) was 6.6±6.6 years. The most common comorbidities were diabetes mellitus (40.2%) and hypertension (60.3%). Poor sleep quality (PSQI>5) was identified in 55% of the patients. There were no statistically significant differences between the good and poor sleep quality groups in terms of age, sex, dialysis duration, or distribution of comorbidities (p>0.05) (**Table 1**).

Table 1. Demographic and clinical characteristics of patients

Variable	Total	Good sleep quality (PSQI≤5)	Poor sleep quality (PSQI>5)	p-value
Age (mean±SD)	59.1±13.5	59.1±12.9	59.0±14.0	0.968
Female (%)	38.9%	41.9%	36.7%	0.463
Male (%)	61.1%	58.1%	63.3%	0.463
CKD duration (years)	6.6±6.6	6.5±6.1	6.7±7.0	0.875
Diabetes-yes (%)	40.2%	44.7%	36.7%	0.259
Hypertension-yes (%)	60.3%	60.0%	60.6%	0.938
COPD-yes (%)	8.8%	5.9%	11.0%	0.210
Heart failure-yes (%)	10.3%	9.4%	11.0%	0.717
Coronary artery disease yes (%)	8.8%	5.9%	11.0%	0.210

PSQI: Pittsburgh Sleep Quality Index, CKD: Chronic kidney disease, COPD: Chronic obstructive pulmonary disease

In contrast, some biochemical and hematological parameters differed significantly between groups. Mean hemoglobin level was 11.1±1.4 g/dl. It was significantly higher in patients with good sleep quality compared to those with poor sleep quality (11.8±1.0 vs. 10.4±1.4 g/dl; p<0.001). Leukocyte count was significantly higher in the poor sleep group (7.82±2.51 vs. 7.02±2.21; p=0.021). Serum CRP level was markedly higher in the poor sleep group (30.34±37.83 mg/L vs. 9.06±11.86 mg/L; p<0.001). CRP/Albumin ratio was significantly elevated in the poor sleep group (8.95±11.64 vs. 2.53±3.54; p<0.001). Serum albumin was higher in the good sleep group (3.79±0.40 vs. 3.60±0.34 g/dl; p<0.001). HALP score was significantly higher in the good sleep group (77.08±133.47 vs. 42.23±81.02; p=0.0002). Serum phosphorus was significantly higher in the poor sleep group (5.23±1.12 vs. 4.53±1.18 mg/dl; p<0.001). Other variables such as neutrophils, lymphocytes, monocytes, ferritin, and PTH did not show statistically significant differences between groups (p>0.05) (**Table 2**).

Table 2. Comparison of laboratory parameters according to PSQI groups

Variable	Overall mean±SD	Good sleep quality (PSQI≤5)	Poor sleep quality (PSQI>5)	p-value
Hemoglobin[A]	11.101±1.493	11.881±1.084	10.484±1.486	<0.001
Leukocytes[A]	7.464±2.412	7.016±2.215	7.820±2.512	0.021
Neutrophils[B]	5.54±1.96	5.47±1.79	5.57±2.04	0.896
Lymphocytes[B]	2.55±4.46	3.25 ±6.27	2.24±3.32	0.251
Monocytes[B]	0.74±0.25	0.73±0.26	0.74±0.24	0.990
Platelets[A]	205.702±70.77	192.547±60.742	216.081±76.461	0.021
Calcium (mg/dl)[A]	8.776±0.767	8.815±0.742	8.745±0.789	0.53
Phosphorus (mg/dl)[A]	5.150±1.358	4.532±1.187	5.228±1.116	<0.001
ALT (U/L)[B]	12.374±9.876	12.035±6.083	12.642±12.080	0.413
Ferritin (ng/ml)[B]	591.569±354.600	586.419±317.205	595.633±382.935	0.733
PTH (pg/ml)[B]	284.043±337.559	268.061±269.738	296.506±382.986	0.92
CRP (mg/L)[B]	20.955±31.152	9.064±11.859	30.336±37.831	<0.001
Albumin (g/dl)[A]	3.686 ±0.379	3.794±0.400	3.600±0.339	<0.001
CRP/albumin[A]	6.117±9.548	2.526±3.540	8.950±11.642	<0.001
HALP score [B]	57.60 ±108.45	77.08±133.47	42.23±81.02	<0.001

A: Independent samples t-test, B: Mann-Whitney U test, PSQI: Pittsburgh Sleep Quality Index, ALT: Alanine aminotranspherase, PTH: Parathyroid hormone, CRP: C-reactive protein, HALP: Hemoglobin albumin lymphocyte platelet

Spearman correlation analysis, demonstrated that leukocyte count (r=0.209, p=0.003), platelet count (r=0.188, p=0.009), CRP (r=0.565, p<0.001), and CRP/albumin ratio (r=0.567, p<0.001) were positively correlated with poor sleep quality. In contrast, hemoglobin (r=–0.618, p<0.001), serum albumin (r=–0.346, p<0.001), and HALP score (r=–0.308, p<0.001) showed significant negative correlations, indicating that lower values of these parameters were associated with poorer sleep (**Table 3**).

Table 3. Spearman correlation

Parameter	Spearman r	p-value
Leukocytes	0.209	0.0034
Platelets	0.188	0.0086
Lymphocytes	0.02	0.78
Monocytes	0.017	0.8106
Neutrophils	0.043	0.5484
Hemoglobin	-0.618	<0.001
Calcium (mg/dl)	-0.061	0.3937
Phosphorus (mg/dl)	-0.039	0.587
ALT (U/L)	-0.098	0.1744
Albumin (g/dl)	-0.346	<0.001
Ferritin (ng/ml)	-0.003	0.9723
CRP (mg/L)	0.565	<0.001
PTH (pg/ml)	-0.029	0.6925
KT/V	-0.11	0.1254
CRP/albumin	0.567	<0.001
HALP score	-0.308	<0.001

ALT: Alanine aminotranspherase, CRP: C-reactive protein, PTH: Parathyroid hormone, HALP: Hemoglobin albumin lymphocyte platelet

Multivariable logistic regression analysis identified elevated CRP and low hemoglobin as independent predictors of poor sleep quality (PSQI>5). Each 1 mg/L increase in CRP was associated with a 3.2% increase in the likelihood of poor sleep (OR 1.032; 95% CI 1.001–1.064; p=0.043), while each 1 g/dl increase in hemoglobin reduced the risk by 52% (OR 0.481; 95% CI 0.350–0.660; p<0.001). Albumin (OR 0.489; p=0.188), HALP score (OR 0.998; p=0.367), phosphorus (OR 1.052; p=0.704), and leukocyte count (OR 1.146; p=0.075) were not statistically significant independent predictors in the multivariable model (**Table 4**).

Table 4. Logistic regression analyses of factors associated with poor sleep quality

	Univariate			Multivariable		
Parameter	OR	95% CI	p-value	OR	95% CI	p-value
Leukocyte	1.16	1.02–1.318	0.0233	1.146	0.986–1.331	0.075
Hemoglobin	0.418	0.311–0.562	0.0	0.481	0.35–0.66	<0.001
Phosphorus	0.971	0.788–1.196	0.7786	1.052	0.809–1.368	0.704
Albumin	0.222	0.093–0.527	0.0007	0.489	0.168–1.419	0.188
CRP	1.061	1.03–1.093	0.0001	1.032	1.001–1.064	0.043
HALP score	1.012	0.974–1.051	0.5533	0.998	0.621–1.192	0.367

OR: Odds ratio, CI: Confidence interval, CRP: C-reactive protein, HALP: Hemoglobin albumin lymphocyte platelet

DISCUSSION

This study revealed that more than half of the hemodialysis patients had poor sleep quality, which was significantly associated with increased inflammatory burden. In the univariate analysis, patients with poor sleep quality had significantly higher levels of CRP, leukocyte count, and CRP/albumin ratio, while hemoglobin, albumin, and HALP scores were significantly lower.

The HALP score is a composite marker that can be easily calculated from hemoglobin, albumin, lymphocyte, and platelet values. It has been widely studied in various patient

populationsespecially cancer patients as a reliable and cost-effective biomarker for predicting prognosis, survival, and treatment response.[12] In recent years, the prognostic role of the HALP score has also been considered in hemodialysis patients. In a large multicenter study by Zhang et al.[13] (n=4,796), a high HALP score was significantly associated with reduced all-cause and cardiovascular mortality in hemodialysis patients. Similarly, Babovic et al.[14] demonstrated a significant positive association between HALP score and quality of life in hemodialysis patients.

In light of this literature, the HALP score was included in our study due to its accessibility and its ability to reflect both inflammation and nutritional status holistically. However, in our patient group, only elevated CRP (OR 1.032; 95% CI 1.001–1.064; p=0.043) and low hemoglobin (OR 0.481; 95% CI 0.350–0.660; p<0.001) emerged as independent predictors of poor sleep quality in the multivariable logistic regression model. The HALP score did not demonstrate independent predictive value. This suggests that while the HALP score is a valuable biomarker for assessing general health and prognosis in clinical practice, it may not be sufficient on its own to predict complex and dynamic clinical outcomes such as sleep quality.

Furthermore, our findings highlight that sleep quality is not merely a subjective symptom but is strongly linked to the balance of inflammation and anemia in hemodialysis patients.

The high prevalence of sleep disorders among hemodialysis patients has been well established in the literature across various regions.[15] Reported prevalence of poor sleep quality ranges from 35% to 85% in different studies. For instance, a study from Saudi Arabia found poor sleep quality in 36.4% of patients and identified associations with older age (>50 years), male sex, marital status, low physical activity, and multiple comorbidities.[16] Another cohort with a mean age of 54 years reported a prevalence of sleep disorders as high as 86.6%.[17] In our study, 55% of patients had PSQI scores >5, but no statistically significant differences were found between sleep quality groups in terms of age, sex, or comorbidities.

This study demonstrated that sleep quality is not solely a subjective experience, but is also meaningfully associated with biochemical and inflammatory parameters. In this respect, our findings support that sleep quality is a holistic health indicator in hemodialysis patients. In particular, CRP levels were significantly higher, and hemoglobin levels were markedly lower in individuals with poor sleep quality.

These findings align with many previous studies. For example, Chiu et al.[18] reported significantly higher CRP and IL-1β levels in hemodialysis patients with poor sleep, along with lower hemoglobin and phosphorus levels. In our study, elevated phosphorus levels were associated with poor sleep quality, potentially due to dietary habits or poor adherence to phosphate binder therapy in our patient group. In contrast, the Chiu[18] study may have observed phosphorus deficiency linked to malnutrition. Taraz et al.[19] found that IL-10, an anti-inflammatory cytokine, was significantly lower in patients with poor sleep quality and was accompanied by higher triglyceride levels. Razeghi et al.[9] reported that high CRP levels and advanced age were significantly associated with sleep disorders in hemodialysis patients. Similarly, studies conducted in different populations have shown that low albumin, high comorbidity burden, advanced age, and elevated inflammatory markers negatively impact sleep quality.[16] Another study reported that patients with at least one sleep disorder had significantly more frequent CRP levels ≥3.8 μg/ml, and that advanced age was also associated with poor sleep (p=0.004 and p=0.006).[9]

Compared with general population studies, the correlations observed in our dialysis cohort are markedly stronger. In a large study of 12.614 Chinese adults, individuals with anemia were more likely to experience insomnia compared with those without anemia; however, the strength of this association was weaker than that observed in our hemodialysis cohort.[20] Similarly, in a study of textile workers in Iran, both ferritin and hemoglobin levels were found to be significantly and inversely associated with sleep quality scores.[21] Finally, an analysis of U.S. adults demonstrated that higher CRP levels were significantly related to sleep disturbances, although the magnitude of this association was lower than in our cohort.[22]

The common thread among these studies is the recognition that systemic inflammation can contribute to the deterioration of sleep quality, and that sleep-related biomarkers (such as CRP, IL-1β, triglycerides, hemoglobin, albumin, and phosphorus) exhibit significant changes. In line with these findings, our study also identified CRP and hemoglobin levels as independent predictors of sleep quality in hemodialysis patients. These results reinforce the impact of inflammation on sleep regulation. Thus, identifying sleep disturbances and implementing therapeutic strategies and lifestyle interventions targeting inflammation have the potential to improve sleep quality. Future prospective studies in this field are clearly needed.

Limitations

This study has several limitations: it is a single-center, cross-sectional study, so causal relationships cannot be definitively established. Objective sleep assessment methods (e.g., polysomnography) were not used. Additionally, other factors known to affect sleep quality, such as depression, restless leg syndrome, and pruritus, were not assessed.

CONCLUSION

This study demonstrated that poor sleep quality in hemodialysis patients is independently associated with elevated CRP levels and low hemoglobin levels. In the univariate analysis, poor sleep was also significantly associated with classical inflammatory markers such as CRP, leukocyte count, and CRP/albumin ratio, as well as the HALP score. These findings indicate that systemic inflammation and anemia play determinant roles in the regulation of sleep patterns. This study showed that although composite biomarkers like HALP may not function as independent predictors in multivariable models, they may serve as useful screening tools at the univariate level; proving themselves as valuable tools while assessing sleep quality. Additionally, it was seen that in dialysis patients although anemia and inflammation

are signficant predictors of sleep quality, as shown in previous studies, phosphorus is also an important element to monitor.

In clinical practice, systematic evaluation of sleep quality in hemodialysis patients and its monitoring through biomarkers may be valuable for early detection of inflammation. Therefore, it is necessary to develop sleep-targeted intervention strategies that both promote a multidisciplinary approach and address inflammation. Furthermore, prospective studies targeting increasing sleep quality could possible show a lessening of inflammation; proving that the relationship between sleep and imflammation is bidirectional. Further prospective studies are warranted in this area.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study has been approved by the Scientific Researches Evaluation and Ethics Committee of Ankara Etlik City Hospital (Date: 18.12.2024, Decision No: 2024-1230).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Seyahi N, Kocyigit İ, Eren N, et al. Current status of kidney replacement therapy in Turkiye: a summary of 2022 Turkish Society of Nephrology registry report. *Turk J Nephrol*. 2024;33(2):134-139. doi:10.5152/turkjnephrol.2025.24978
2. Pawar YS, Gattani VS, Chaudhari KS, Chheda B, Vankudre AJ. Impact of hemodialysis on sleep disorders in patients with end-stage renal disease in a tertiary care academic hospital. *Cureus*. 2023;15(8):e44416. doi:10.7759/cureus.44416
3. Ezzat H, Mohab A. Prevalence of sleep disorders among ESRD patients. *Ren Fail*. 2015;37(6):1013-1019. doi:10.3109/0886022X.2015.1044401
4. Iliescu EA, Coo H, McMurray MH, et al. Quality of sleep and health-related quality of life in haemodialysis patients. *Nephrol Dial Transplant*. 2003;18(1):126-132. doi:10.1093/ndt/18.1.126
5. Belluardo G, Sessa C, Morale W. The alterations of sleep and frontal functions in chronic hemodialysis: pathogenesis and therapeutic perspectives. *Behav Brain Res*. 2025;478:115337. doi:10.1016/j.bbr.2024.115337
6. Turgay G, Tutal E, Sezer S. Restless legs syndrome, sleep quality and fatigue in hemodialysis patients. *Turk Neph Dial Transplant*. 2018;27(2):196-203.
7. Elder SJ, Pisoni RL, Akizawa T, et al. Sleep quality predicts quality of life and mortality risk in haemodialysis patients: results from the dialysis outcomes and practice patterns study (DOPPS). *Nephrol Dial Transplant*. 2008;23(3):998-1004. doi:10.1093/ndt/gfm630
8. Russcher M, Chaves I, Lech K, et al. An observational study on disturbed peripheral circadian rhythms in hemodialysis patients. *Chronobiol Int*. 2015;32(6):848-857. doi:10.3109/07420528.2015.1048868
9. Razeghi E, Sahraian MA, Heidari R, Bagherzadeh M. Association of inflammatory biomarkers with sleep disorders in hemodialysis patients. *Acta Neurol Belg*. 2012;112(1):45-49. doi:10.1007/s13760-012-0003-7
10. 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(2):193-213. doi:10.1016/0165-1781(89)90047-4
11. Ağargün MY, Kara H, Anlar Ö. Pittsburgh uyku kalitesi indeksinin geçerligi ve güvenirligi. *Turk Psikiyatri Dergisi*. 1996;7:107-115.
12. Farag CM, Antar R, Akosman S, Ng M, Whalen MJ. What is hemoglobin, albumin, lymphocyte, platelet (HALP) score? A comprehensive literature review of HALP's prognostic ability in different cancer types. *Oncotarget*. 2023;14:153-172. doi:10.18632/oncotarget.28367
13. Zhang F, Li L, Shi T, Liu Y, Xie J, Yu L. The hemoglobin, albumin, lymphocyte, and platelet (HALP) is a potent indicator for the prognosis in hemodialysis patients. *Medicine (Baltimore)*. 2023;102(19):e33650. doi:10.1097/MD.0000000000033650
14. Babovic B, Belada Babovic N, Tomovic F, et al. The importance of biochemical parameters, immunonutritional status, and social support for quality of life in chronic hemodialysis patients. *Medicina (Kaunas)*. 2024;60(11):1751. doi:10.3390/medicina60111751
15. Merlino G, Piani A, Dolso P, et al. Sleep disorders in patients with end-stage renal disease undergoing dialysis therapy. *Nephrol Dial Transplant*. 2006;21(1):184-190. doi:10.1093/ndt/gfi144
16. Alshammari B, Alkubati SA, Pasay-An E, et al. Sleep quality and its affecting factors among hemodialysis patients: a multicenter cross-sectional study. *Healthcare (Basel)*. 2023;11(18):2536. doi:10.3390/healthcare11182536
17. Masoumi M, Naini AE, Aghaghazvini R, Amra B, Gholamrezaei A. Sleep quality in patients on maintenance hemodialysis and peritoneal dialysis. *Int J Prev Med*. 2013;4(2):165-172.
18. Chiu YL, Chuang YF, Fang KC, et al. Higher systemic inflammation is associated with poorer sleep quality in stable haemodialysis patients. *Nephrol Dial Transplant*. 2009;24(1):247-251. doi:10.1093/ndt/gfn439
19. Taraz M, Khatami MR, Hajiseyedjavadi M, et al. Association between antiinflammatory cytokine, IL-10, and sleep quality in patients on maintenance hemodialysis. *Hemodial Int*. 2013;17(3):382-390. doi:10.1111/hdi.12035
20. Neumann SN, Li JJ, Yuan XD, et al. Anemia and insomnia: a cross-sectional study and meta-analysis. *Chin Med J (Engl)*. 2020;134(6):675-681. doi:10.1097/CM9.0000000000001306
21. Ghahremanfard F, Semnani MR, Mirmohammadkhani M, et al. The relationship between iron deficiency anemia with restless leg syndrome and sleep quality in workers working in a textile factory in Iran: a cross-sectional study. *Middle East Curr Psychiatry*. 2023;30(1):23. doi:10.1186/s43045-023-00294-5
22. Liu R, Liu X, Zee PC, et al. Association between sleep quality and C-reactive protein: results from national health and nutrition examination survey, 2005-2008. *PLoS ONE* 2014;9(3):e92607. doi:10.1371/journal.pone.0092607

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1747735

Anatolian Curr Med J. 2025;7(5):618-622

Evaluation of local anesthesia knowledge levels of specialist/PhD dentists in Turkiye

Erkan Feslihan[1], Hilal Gülgezen Aydın[2], Tuğçe Dönmezer[3], Volkan Kaplan[1]

[1]*Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, Tekirdağ Namık Kemal University, Tekirdağ, Turkiye*

[2]*Department of Prosthodontics, Faculty of Dentistry, Tekirdağ Namık Kemal University, Tekirdağ, Turkiye*

[3]*Department of Periodontology, Faculty of Dentistry, Tekirdağ Namık Kemal University, Tekirdağ, Turkiye*

Cite this article as: Feslihan E, Gülgezen Aydın H, Dönmezer T, Kaplan V. Evaluation of local anesthesia knowledge levels of specialist/PhD dentists in Turkiye. *Anatolian Curr Med J.* 2025;7(5):618-622.

Received: 21.07.2025 ◆ **Accepted**: 25.08.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: This study aims to evaluate the knowledge levels of specialist/PhD dentists about local anesthetic agents, application techniques, and complications related to local anesthesia.

Methods: An electronic survey was distributed to specialized/PhD dentists employed in 8 different departments of dentistry. The survey consisted of 25 questions about knowledge of local anesthesia. Knowledge level was determined based on the number of correct answers. In the statistical evaluation, descriptive statistics for continuous variables were reported as mean, standard deviation, median, minimum, and maximum values. Categorical variables were expressed as frequency and percentage. For continuous variables, the data was assessed using the Shapiro-Wilk test. For independent group comparisons involving more than two groups, the Kruskal-Wallis variance analysis test was employed. In subgroup comparisons, the Bonferroni-adjusted Mann-Whitney U test was utilized. For comparisons between two independent groups, the Mann-Whitney U test was applied. The statistical significance level was set at 0.05.

Results: This study involved 200 participants, comprising 35.5% males and 64.5% females, with an average age of 35.91 years. Participants were categorized based on their years of service, with the majority having 5-10 years of experience. A significant portion (60.5%) had received specialized training, and 87.0% were employed in the public sector. Knowledge levels regarding local anesthesia varied among different dental specialties, with oral and maxillofacial surgery showing the highest proficiency. While no significant differences in knowledge were found based on gender, years of experience, or educational degree, a notable distinction emerged among those with PhDs or specializations, particularly favoring oral and maxillofacial surgery. Conversely, prosthodontics specialists exhibited lower knowledge levels compared to several other fields.

Conclusion: Dentists' knowledge of local anesthesia varies across specialty/PhD fields, with oral and maxillofacial surgeons having the highest scores and Prosthodontists having the lowest, highlighting the need for continuous training.

Keywords: Local anesthesia, knowledge, survey, specialization, PhD

INTRODUCTION

Local anesthesia is defined as the temporary loss of sensation in a specific area of the body by suppressing the stimulation in the nerve endings or by blocking the conduction process in peripheral nerves.[1,2] Local anesthesia can be provided by several mechanisms, including electrical nerve stimulation, cryotherapy, and low-level laser, in addition to traditional local anesthetic drugs.[3] The most common method for pain management in dental practice is the use of reversible local anesthetic agents.[4]

Local anesthesia in dentistry is usually applied topically or by injection.[5] Topical anesthetics are agents that block the conduction of nerve impulses in superficial nerves, causing a loss of sensation in the mucosa or skin. They are available in various forms, such as spray, cream, and gel, and are recommended for use in mild clinical conditions, including aphthous ulcer treatment and reducing needle pain during injection.[6] Injectable local anesthetics are divided into amide and ester groups based on their chemical structures, enabling many dental procedures, such as tooth extraction, pulp treatments, restorative procedures, periodontal and surgical interventions, to be performed safely and comfortably (**Table 1**).[7,8] However, other components, such as vasoconstrictors found in the local anesthetic solutions, are of great importance during local anesthetic application.[9] Vasoconstrictors such as epinephrine cause constriction in the surrounding vascular structures, limiting the absorption of the local anesthetic agent into the bloodstream and thus prolonging its duration of action.[10] When calculating the maximum dose for medically

Corresponding Author: Volkan Kaplan, dr.volkankaplan61@gmail.com

compromised patients and children, factors such as age and weight, as well as the vasoconstrictor content of the anesthetic solution, should be considered. Repeated injections after inadequate anesthesia may cause neurological and cardiac complaints due to local drug concentration reaching high levels in a short time.[11]

Table 1. Classification of local anesthetic agents based on chemical structure

Ester group	Amide group
Cocaine	Lidocaine
Chloroprocaine	Mepivacaine
Procaine	Prilocaine
Tetracaine	Bupivacaine
Benzocaine	Etidocaine
	Articaine

Local anesthetics are safe and effective drugs, but they can cause several local and systemic complications that dentists should be aware of.[12] Local complications include paresthesia, pain, trismus, infection, and ocular complications such as double vision or temporary blindness. Rare systemic complications, including allergic reactions, systemic toxicity, and methemoglobinemia, may lead to life-threatening conditions.[13-15]

Recently available local anesthetic agents and anesthesia delivery equipment provide dentists with numerous options to manage pain associated with dental procedures effectively.[16] To achieve successful local anesthesia, the dentist should possess precise knowledge of the drugs, relevant neuroanatomy, anesthesia techniques, and equipment, as well as the ability to diagnose and treat possible complications promptly.[17]

This study aims to evaluate the knowledge levels of specialist/PhD dentists about local anesthetic agents, application techniques, and complications related to local anesthesia. Thus, if there are any deficiencies in the local anesthesia knowledge level of specialist/PhD dentists, this will contribute to the studies on eliminating these deficiencies. The hypothesis of this study is that there is no difference in the level of local anesthesia knowledge among specialist/PhD dentists.

METHODS

The study protocol was approved by the Non-interventional Researches Ethics Committee of Tekirdağ Namık Kemal University (Date: 28.01.2025, Decision No: 2025.21.01.21). Throughout the study, ethical standards for research and publication were strictly followed. Participants completed the research surveys between March 1, 2025, and March 14, 2025. All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

This study was conducted through electronically distributed surveys targeting dentists actively practicing in clinics affiliated with public or private universities, as well as in public or private oral and dental health centers and hospitals within our country. Participants were required to have completed a PhD or specialization training in various dental disciplines. Dentists who did not possess a PhD or specialization, or those who were not treating patients at the time of the survey, were excluded from the study. Data were collected through responses to a survey (Google Forms) accessed via an email link. The survey began with a statement outlining the purpose of the study and a question asking whether participants agreed to participate. Participants who responded to this question and completed the survey were considered to have provided voluntary informed consent.

A power analysis was conducted to determine the appropriate sample size. It was found that with an alpha level of 0.05 and a power set at 80%, a small effect size (effect size=0.20) requires a sample size of 368, while a medium effect size (effect size=0.50) requires 72 participants.[18] In this study, eight groups were utilized: group 1 (oral and maxillofacial surgery), group 2 (oral and maxillofacial radiology), group 3 (endodontics), group 4 (orthodontics), group 5 (pedodontics), group 6 (periodontology), group 7 (prosthodontics), and group 8 (restorative dentistry). Consequently, the sample size for the eight groups was accepted as 200, corresponding to an effect size of 0.27 (n=25). A total of 200 individuals participated, with 25 participants from each group. The completed questionnaires collected demographic information, including age, gender, education level, work experience, and training status, as well as responses to 25 questions assessing knowledge of local anesthesia. The survey questions were selected from the questions already prepared by the authors of the textbook written by Koçak et al.[19]

Knowledge level evaluations were determined based on the number of correct answers given by individuals in the surveys. The classifications were as follows: weak (fewer than 13 correct answers out of 25 questions), moderate (13-18 correct answers out of 25 questions), and good (more than 18 correct answers out of 25 questions).

Statistical Analysis

In this study, the level of knowledge was examined as the primary characteristic. Due to the nature of online surveys, where questions and answers are predetermined and survey data is not recorded in the system when responses are missing, there was no loss of respondents. For the emphasized characteristics, descriptive statistics for continuous variables were reported as mean, standard deviation, median, minimum, and maximum values. Categorical variables were expressed as frequency and percentage. For continuous variables, the normality of the data was assessed using the Shapiro-Wilk test. Since normality was not met for independent group comparisons involving more than two groups based on categorical variables, the Kruskal-Wallis variance analysis test was employed. In subgroup comparisons, the Bonferroni-adjusted Mann-Whitney U test was utilized. For comparisons between two independent groups, the Mann-Whitney U test was applied. The statistical significance level was set at 0.05, and the computations were performed using the SPSS statistical Software package (version 21).

RESULTS

Among the 200 individuals who participated in the study, 71 (35.5%) were male and 129 (64.5%) were female, with an

average age of 35.91±6.20 years. Each group consisted of 25 individuals. When evaluated according to their years of service, the proportions of participants with 0-2 years, 3-4 years, 5-10 years, 11-15 years, and more than 16 years of service were 21.0%, 23.0%, 32.0%, 14.5%, and 9.5%, respectively. Additionally, 60.5% of the survey participants had received specialized training, and 87.0% were employed in the public sector (**Table 2**).

Table 2. Descriptive statics

Age	Mean±SD	35.91±6.20
	Median	35.00
	Min-max	27-58
		Frequency (%)
Gender	Male	71 (35.50)
	Female	129 (64.50)
Education	PhD	79 (39.50)
	Specialty	121 (60.50)
Area of PhD/specialty	Oral and maxillofacial surgery	25 (12.5)
	Oral and maxillofacial radiology	25 (12.5)
	Endodontics	25 (12.5)
	Orthodontics	25 (12.5)
	Pedodontics	25 (12.5)
	Periodontology	25 (12.5)
	Prosthodontics	25 (12.5)
	Restorative dentistry	25 (12.5)
Work experience (years)	0-2	42 (21.00)
	3-4	46 (23.00)
	5-10	64 (32.00)
	11-15 16+	29 (14.50) 19 (9.50)
Workplace	Public employee	174 (87.00)
	Private practice	26 (13.00)

SD: Standard deviation, %: percent, Min: Minimum, Max: Maximum

When evaluating the responses of the groups examined in the study, the following groups demonstrated a good level of knowledge regarding local anesthesia: oral and maxillofacial surgery (24 individuals), pedodontics (19 individuals), orthodontics (16 individuals), oral and maxillofacial radiology (15 individuals), restorative dentistry (15 individuals), periodontology (14 individuals), endodontics (13 individuals), and prosthodontics (9 individuals). While individuals with medium level of local anesthesia knowledge were present in all groups, those with a weak level of local anesthesia knowledge were found only in the prosthodontics (4 individuals) and orthodontics (1 individual) groups (**Table 3**). The most frequently incorrect response given by the participants (81.5%) was to the question about the vitamin that plays the most significant role in repairing nerve damage whereas the question that 99% of the participants answered correctly, was about the branches of the cranial nerve anesthetized for pain control in the oral cavity and facial region.

Table 3. The groups' levels of knowledge regarding local anesthesia

Area of PhD/specialty	Number of participants		
	Good (%)	Medium (%)	Weak (%)
Oral and maxillofacial surgery	24 (96)	1 (4)	0 (0)
Oral and maxillofacial radiology	15 (60)	10 (40)	0 (0)
Endodontics	13 (52)	12 (48)	0 (0)
Orthodontics	16 (64)	8 (32)	1 (4)
Pedodontics	19 (76)	6 (24)	0 (0)
Periodontology	14 (56)	11 (44)	0 (0)
Prosthodontics	9 (36)	12 (48)	4 (16)
Restorative dentistry	15 (60)	10 (40)	0 (0)

Weak: Less than 13 correct answers out of 25 questions, Medium: 13-18 correct answers out of 25 questions, Good: More than 18 correct answers out of 25 questions, %: percent

There was no significant difference in the evaluation of knowledge levels regarding local anesthesia among the study participants based on gender, years of work experience, or educational degree ($p>0.05$). However, when assessing the knowledge levels of local anesthesia among individuals according to their PhD or specialization fields, a significant difference was observed ($p=0.000$) (**Table 4**). Specifically, participants with a PhD or specialization in oral and maxillofacial surgery demonstrated higher levels of knowledge in local anesthesia compared to those in the fields of endodontics, orthodontics, periodontology, prosthodontics, and restorative dentistry. Furthermore, individuals with a PhD or specialization in prosthodontics exhibited lower levels of knowledge about local anesthesia compared to those in the fields of oral and maxillofacial radiology, pedodontics, periodontology, and restorative dentistry.

DISCUSSION

In this survey-based, cross-sectional study, the knowledge levels of dentists from different fields regarding local anesthetics were measured. The survey was distributed through online channels and via email. When the level of knowledge among the fields of specialization was evaluated, the dentists in the field of oral and maxillofacial surgery were found to be statistically significantly more successful with more correct answers compared to other fields (21.68±1.95). These results indicate that the level of knowledge of Oral and maxillofacial surgeons regarding nerve anatomy, local anesthetic agents, and pain control is higher compared to other fields. On the other hand, prosthodontists showed significantly lower success rates compared to other fields (16.24±4.06). Similarly Tadin et al.[17] reported that a higher knowledge level was associated with specialization in oral surgery while a lack of confidence in using various local anesthetic techniques was also associated with specialization in orthodontics. It was also indicated that the knowledge level of local anesthesia were higher in females.[17] In our study there were no significant difference between participants regarding to gender and years of work experience. In consistent with our results Bani-Hani et al.[20] demonstarated that gender and level of experience did not significantly influence specialists' practice or their knowledge of local anesthesia.

Table 4. Comparison of the total number of correct answers in the level of knowledge about local anesthesia according to gender, years of work experience, educational degree, and position

		Number of participants	Mean level of knowledge±SD (n)	Min-max (n)	Median (n)	p
Gender	Male	71	18.80±3.60	6-25	19.00	0.849
	Female	129	19.02±2.70	11-24	19.00	
Work experience (years)	0-2	42	18.52±2.82	12-24	19.00	0.523
	3-4	46	19.24±3.20	6-24	19.50	
	5-10	64	19.33±2.95	11-24	20.00	
	11-15	29	18.66±3.06	11-25	19.00	
	16+	19	18.26±3.44	12-23	20.00	
Education	PhD	79	18.39±3.68	6-24	19.00	0.141
	Specialty	121	19.30±2.49	13-25	20.00	
Area of PhD/ specialty	Oral and maxillofacial surgery	25	21.68±1.95	17-25	22.00	0.000
	Oral and maxillofacial radiology	25	19.56±2.63[f1]	15-24	19.00	
	Endodontics	25	18.48±2.00[a1]	14-23	19.00	
	Orthodontics	25	18.52±3.29[b1]	12-24	19.00	
	Pedodontics	25	19.48±2.08[c1,g1]	16-24	20.00	
	Periodontology	25	18.76±2.65[h1]	14-24	19.00	
	Prosthodontics	25	16.24±4.06[d1]	6-23	16.00	
	Restorative dentistry	25	18.80±2.61[e1,i1]	13-23	19.00	

[a] Oral and maxillofacial surgery-endodontics, [b] Oral and maxillofacial surgery-orthodontics, [c] Oral and maxillofacial surgery-pedodontics, [d] Oral and maxillofacial surgery-prosthodontics, [e] Oral and maxillofacial surgery-restorative dentistry, [f] Prosthodontics-oral and maxillofacial radiology, [g] Prosthodontics-pedodontics, [h] Prosthodontics-periodontology, [i] Prosthodontics-Restorative dentistry, [1]p<0.05, SD: Standard deviation, n: Number of correctly answered questions, Min: Minimum, Max: Maximum

In the recent study the highest number of incorrect answers to the question about treating the nerve damage occurred as a complication of local anesthesia may be attributed to a decline in dentists' knowledge, as vitamin B complex is commonly prescribed for cases of paresthesia. However, it was found that participants answered the fundamental knowledge question related to the anesthesia applied to the branches of the N. Trigeminus for pain management during dental procedures correctly at a high rate. Accordingly, the need for different levels of anesthesia in the daily clinical practice of physicians in different fields may have ensured the continuity of the theoretical knowledge they acquired. Additionally, it is thought that these differences may also stem from the variations in course content provided during specialization/ PhD training. There is no standard regarding the content of courses offered in PhD and specialization programs at universities in the specialized fields of dentistry. Therefore, it is not possible to directly attribute the intergroup differences to the educational content. However, it is recommended to update the training programs for dentists in the field of Prosthodontics and to place greater emphasis on topics such as nerve anatomy, local anesthetic knowledge, and pain control. Additionally, organizing interactive training programs that encourage the sharing of knowledge and experience among dentists working in this field may be beneficial.

In this study, a weak level of knowledge was detected among all participating dentists, with a rate of 2.5%. Although this rate is low, it is unacceptable because it includes the application of local anesthesia, which is a fundamental practice for human health. As suggested by Kaira and Dabral,[11] it is recommended that dentists receive further education and keep their knowledge up to date in critical applications.

In this study, the correct answer rate of all participants to the knowledge questions was 75.7%. In a study evaluating the knowledge level of physician research assistants regarding local anesthetics and their toxicity, the correct answer rate among participants was reported as 74.4%, similar to the rate observed in this study.[15] In another study evaluating the knowledge level of dentists regarding safe local anesthetic practices, the correct answer rate was reported to be 40.4%.[21] The reason for the lower correct answer rate compared to the rate reported in this study may be that more topic-specific questions were asked about local anesthetic systemic toxicity (LAST).

In a study examining the level of knowledge regarding local anesthetic doses, a comparison was made between specialists and dentists; however, the level of knowledge across different specialties was not compared.[11] To the authors' knowledge, no study in the current literature has evaluated the differences in general knowledge level of local anesthesia among the specialties of dentistry. Future studies should concentrate on assessing the knowledge and practice levels of dentists across various specialties. This approach is expected to contribute to the development of postgraduate education programs and enhance treatment success rates.

According to the results of this study, the hypothesis that there is no difference in the level of local anesthesia knowledge among specialist/PhD dentists was rejected because a difference was detected between the groups.

Limitations

There are a few issues regarding the limitations of the present study. The surveys were not conducted face-to-face and distributing the survey solely through online channels and

e-mail may have led to sampling bias due to the exclusion of professionals who do not use these platforms. In addition, a relatively small sample size could limit the extent of our findings to be extrapolated in a larger population.

CONCLUSION

There are notable differences in the knowledge of local anesthesia among dentists across various specialty/PhD fields. Oral and maxillofacial surgeons achieved the highest scores, while prosthodontists obtained the lowest scores. Although the percentage of weak level of knowledge is low (2.5%), this is unacceptable because it includes the application of local anesthesia, a critical practice in dental health. Therefore, graduate dentists should be provided with online training through associations/unions or face-to-face training at congresses/symposiums to ensure they keep their knowledge up-to-date. Studies involving more participants are recommended.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study protocol was approved by the Non-Interventional Researches Ethics Committee of Tekirdağ Namık Kemal University (Date: 28.01.2025, Decision No: 2025.21.01.21).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

Acknowledgement

The statistical analysis of this study was performed by Assoc. Prof. Birol TOPÇU (Tekirdağ Namık Kemal University, Faculty of Medicine, Department of Biostatistics).

REFERENCES

1. Ogle OE, Mahjoubi G. Local anesthesia: agents, techniques, and complications. *Dental Clinics*. 2012;56(1):133-148. doi:10.1016/j.cden.2011.08.003
2. Ghafoor H, Haroon S, Atique S, et al. Neurological complications of local anesthesia in dentistry: a review. *Cureus*. 2023;15(12):e50790. doi:10.7759/cureus.50790
3. Patel BJ, Surana P, Patel KJ, Patel B. Recent advances in local anesthesia: a review of literature. *Cureus*. 2023;15(3):e36291. doi:10.7759/cureus.36291
4. Bahl R. Local anesthesia in dentistry. *Anesth Prog*. 2004;51(4):138-142.
5. Pahade A, Bajaj P, Shirbhate U, John HA. Recent modalities in pain control and local anesthesia in dentistry: a narrative review. *Cureus*. 2023;15(11):e48428. doi:10.7759/cureus.48428
6. Franz-Montan M, Ribeiro LNDM, Volpato MC, et al. Recent advances and perspectives in topical oral anesthesia. *Expert Opinion on Drug Delivery*. 2016;14(5):673-684. doi:10.1080/17425247.2016.1227784
7. Haas DA. An update on local anesthetics in dentistry. *J-Canadian Dent Assoc*. 2002;68(9):546-552.
8. Hawkins JM, Moore PA. Local anesthesia: advances in agents and techniques. *Dent Clin*. 2002;46(4):719-732. doi:10.1016/s0011-8532(02)00020-4
9. Decloux D, Ouanounou A. Local anaesthesia in dentistry: a review. *Int Dent J*. 2021;71(2):87-95. doi:10.1111/idj.12615
10. Pinheiro AC, Marques JF, Vieira MS, Branco-De-Almeida LS. Dentists' knowledge regarding signs and symptoms of the systemic toxicity of local anesthetic solutions. *RGO-Rev Gaúch Odontol*. 2015;63:41-46. doi:10.1590/1981-86372015000100062962
11. Kaira LS, Dabral E. A survey to access knowledge and practice among dentists regarding local anestetic dosage in three cities of Uttarakhand. *Eur J Gen Dent*. 2014;3:105-108. doi:10.4103/2278-9626.134832
12. Cummings DR, Yamashita DDR, McAndrews JP. Complications of local anesthesia used in oral and maxillofacial surgery. *Oral Maxillofac Surg Clin North Am*. 2011;23(3):369-377. doi:10.1016/j.coms.2011.04.009
13. Ho JPT, Van Riet TC, Afrian Y, et al. Adverse effects following dental local anesthesia: a literature review. *J Dent Anesth Pain Med*. 2021;21(6):507-525. doi:10.17245/jdapm.2021.21.6.507
14. Mangal SK, Ravindra V. Knowledge and awareness on administration of local anesthesia and its various complications faced by undergraduate and postgraduate dental students-a cross-sectional study. *JETT*. 2022;13(6):103-115. doi:10.47750/jett.2022.13.06.011
15. Karasu D, Yılmaz C, Özgünay ŞE, Dayıoğlu M, Baytar Ç, Korfalı G. Knowledge of the research assistants regarding local anaesthetics and toxicity. *Turk J Anaesthesiol Reanim*. 2016;44:201-205. doi:10.5152/TJAR.2016.53138
16. Saxena P, Gupta SK, Newaskar V, Chandra A. Advances in dental local anesthesia techniques and devices: an update. *Natl J Maxillofac Surg*. 2013;4:19-24. doi:10.4103/0975-5950.117873
17. Tadin A, Aleric K, Jerkovic D, Gavic L. Knowledge, practice and self-reported confidence level of croatian dentists in the use of local anesthesia: a cross-sectional study. *Healthcare*. 2023;11(14):2006. doi:10.3390/healthcare11142006
18. Cohen J. Statistical power analysis for the behavioral sciences (2nd ed.). Routledge; 1988. doi:10.4324/9780203771587
19. Koçak Berberoğlu H, Gürkan Köseoğlu B, Kasapoğlu Ç. Diş Hekimliğinde lokal anestezi. Quintessence; 2007.
20. Bani-Hani T, Al-Fodeh R, Tabnjh A, Leith R. The use of local anesthesia in pediatric dentistry: a survey of specialists' current practices in children and attitudes in relation to articaine. *Int J Dent*. 2024;2024:2468502. doi:10.1155/2024/2468502
21. Bassyoni L, Alshehri M, Sindi M, Elias A, Nassar A. Dental practitioners' knowledge, behavior, and attitude toward safe local anesthesia practice in Jeddah. *J Complement Med Res*. 2022;13:9-15. doi:10.5455/jcmr.2022.13.05.02

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1749095

Anatolian Curr Med J. 2025;7(5):623-629

CD1a expression in HER2-positive breast carcinoma tissues and its prognostic implications

Burcu Özcan[1], Ekrem Yavuz[2]

[1]*Department of Medical Pathology, İstanbul Training and Research Hospital, University of Health Sciences, İstanbul, Turkiye*
[2]*Department of Medical Pathology, İstanbul Faculty of Medicine, İstanbul University, İstanbul, Turkiye*

Cite this article as: Özcan B, Yavuz E. CD1a expression in HER2-positive breast carcinoma tissues and its prognostic implications. *Anatolian Curr Med J.* 2025;7(5):623-629.

Received: 23.07.2025 • **Accepted**: 30.08.2025 • **Published**: 15.09.2025

ABSTRACT

Aims: Dendritic cells are key players in initiating the primary immune response. Many studies have investigated the role of antigen-presenting cells in breast carcinoma, revealing associations between dendritic cell density, their spatial distribution within the tumor, and prognosis. CD1a is a glycosylated type I transmembrane protein that is frequently used as a dendritic cell marker in human tumors. This study aims to detect CD1a expression using immunohistochemical methods in a cohort of HER2-overexpressing breast carcinoma patients. It also analyzes the correlation of CD1a expression with prognostic parameters and overall survival outcomes.

Methods: Patient records of individuals diagnosed with invasive breast carcinoma between 2003 and 2009 were retrieved from the digital pathology archive of İstanbul University İstanbul Faculty of Medicine Pathology Department. From this dataset, 48 cases were selected based on the following inclusion criteria: HER2 positivity confirmed. We applied cd1a immunohistochemically to all of these cases and compared the staining with prognostic data.

Results: Any level of CD1a positivity was considered when evaluating the association between CD1a expression and survival/prognostic factors. No statistically significant relationship was found between CD1a positivity and age, tumor size, lymph node involvement, or recurrence. No correlation was identified between CD1a positivity/negativity and the neoadjuvant status of the tumour. However, a significant association was observed between nuclear grade and CD1a positivity. No statistical correlation was found between 5-year total and disease-free survival and CD1a staining.

Conclusion: Considering that HER2-positive tumors account for only 25–30% of breast carcinomas, our study focused on a particular and homogeneous patient group. A lower survival rate was expected because HER2 positivity is a known poor prognostic factor, and our cohort consisted of non-early-stage cases. Therefore, while a proportional relationship between CD1a positivity and survival was observed in HER2-positive cases, statistical significance was not achieved. Future studies should include larger patient populations or more heterogeneous cohorts in terms of prognostic features to obtain more conclusive results. In our study, CD1a positivity was significantly associated only with higher nuclear grade.

Keywords: CD1a, HER2, breast carcinoma, dendritic cell

INTRODUCTION

Dendritic cells (DCs) are key players in initiating the primary immune response. Originating from the bone marrow, these cells process antigens and present them to T cells. CD34+ myeloid progenitor cells, under the influence of GM-CSF and TNF-α, differentiate into two distinct types of DCs—CD1a+ and CD14+—by day 5 of development. These two subtypes mature into typical DCs by day 12, enter circulation, and reside as immature cells in peripheral organs.[1] They are predominantly located in barrier tissues such as the skin, nasal mucosa, respiratory tract, and gastrointestinal system.[2]

Immature DCs undergo a three-step maturation process involving antigen capture, migration, and antigen presentation to T cells.[1,3,4] Mature DCs are CD83-positive and express high levels of MHC class II molecules, CD40, CD80, and CD86. In contrast, immature DCs are CD1a-positive and show low expression of CD40, CD80, and CD86.[5]

CD1a is a glycosylated type I transmembrane protein that exhibits structural homology with MHC class I and II molecules.[6] It is frequently used as a dendritic cell marker in human tumors. In the lymphoreticular system, CD1a expression is limited to a subset of DCs and some thymic cells.[7]

Numerous studies have demonstrated a correlation between dendritic cell density in tumor tissues and clinical outcomes

Corresponding Author: Burcu Özcan, drburcuozcan@yahoo.com

across various tumor types.[6,8-12] The earlier perception of breast cancer as a non-immunogenic tumor has been challenged. The presence of antibodies specific to breast carcinoma antigens and the expansion of tumor-reactive T-cell clones among tumor-infiltrating lymphocytes support the immunogenic nature of breast cancer.[13]

Many studies have investigated the role of antigen-presenting cells in breast carcinoma, revealing associations between dendritic cell density, their spatial distribution within the tumor, and patient prognosis.[6,13] Since mature DCs are typically observed only in lymphoid organs close to T cells, the appearance of mature DCs in peritumoral regions suggests that tumor-specific immune responses may be initiated in these areas.[13]

However, the origin of tumor-infiltrating DCs remains controversial. Some studies propose that these and Langerhans cells arise from distinct lineages.[14] In contrast, lymphoscintigraphic studies have demonstrated the migration of Langerhans cells from the skin of the breast into breast tissue and subsequently to the axillary lymph nodes.[7] Supporting this view, other reports have shown that CD1a+ Langerhans cells exhibit increased CD1a expression following 24-hour in vitro incubation and lose Birbeck granules—a hallmark of Langerhans cells—during this process.[15]

HER2 is a transmembrane protein which possesses tyrosine kinase activity. HER2 is expressed at low levels in normal mammary epithelium, where only a single copy of the gene exists. In approximately 25–30% of breast carcinomas, HER2 gene amplification—often without changes in chromosome 17 copy number—results in overexpression of the receptor protein. Activation of these overexpressed receptors triggers intracellular signaling cascades that enhance tumor cell survival and proliferation.[16]

This study aims to detect CD1a expression using immunohistochemical methods in a cohort of HER2-overexpressing breast carcinoma patients. It also analyzes the correlation of CD1a expression with prognostic parameters and overall survival (OS) outcomes.

METHODS

Ethics

The study protocol has been approved by the İstanbul University Clinical Researches Ethics Committee (Date: 04.01.2013, Decision No: 2012/1504-1231). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Sample Selection

This retrospective study retrieved patient records of individuals diagnosed with invasive breast carcinoma between 2003 and 2009 from the digital pathology archive of İstanbul University İstanbul Faculty of Medicine Pathology Department. From this dataset, 48 cases were selected based on the following inclusion criteria: HER2 positivity confirmed either as strongly positive (+++) by immunohistochemistry (IHC) or moderately positive (++), with additional verification of HER2 gene amplification via the silver-enhanced in situ hybridization (SISH) technique in five of these cases. Cases that were HER2-negative or scored moderately positive but lacked confirmed positivity by SISH, cases with cell blocks unsuitable for immunohistochemical analysis, and those with unavailable prognostic data were excluded from the study.

Among the 48 tumor samples included in the study, 24 were pre-treatment biopsy specimens, while the remaining 24 were obtained from resected tumor tissues following neoadjuvant therapy. For each case, when both pre-treatment tru-cut biopsy and post-neoadjuvant resection specimens were available, the excised tissue was preferred due to its higher tumor volume and better cellularity. However, in cases where the resection specimen showed marked regression or insufficient viable tumor cells following trastuzumab-based therapy, the pre-treatment biopsy was used for CD1a immunohistochemical evaluation. All cases were reviewed on hematoxylin and eosin (H&E) stained slides, and corresponding paraffin blocks were retrieved from the departmental archives.

All selected patients had stage III breast carcinoma and received neoadjuvant treatment regimens that included trastuzumab. In accordance with the AJCC (8th edition) staging criteria, stage III breast carcinoma encompasses tumors with larger size and/or regional lymph node involvement, including T3N1, T2N2, or T4 lesions, but without distant metastasis (M0). Of the 48 cases included in the study, pathological regression data after neoadjuvant therapy were available for 47 cases. Pathological complete response (pCR) was identified in 12 of these 47 cases (25.5%). Prognostic information was collected from the patient records archived at the Oncology Institute of the İstanbul University İstanbul Faculty of Medicine. Data regarding treatment protocols and recurrence were obtained by reviewing clinical files. Additional survival information was gathered by directly contacting patients or their relatives using details from the institution's digital patient database. Both OS and disease-free survival (DFS) times were recorded. Histologic grading was assessed using the Modified Bloom-Richardson grading system.[17] Due to the use of tru-cut biopsies in some cases, only nuclear grade was assessed based on a modified Bloom-Richardson system for standardization.

Immunhistochemical Staining Method

In accordance with the datasheet recommendations of the primary antibody, non-tumoral thymic tissue was used as the positive control. Additionally, Langerhans cell histiocytosis tissue was included as a tumor-related positive control to ensure staining quality and reproducibility. Negative control staining was performed in parallel by omitting the primary antibody, confirming the specificity of CD1a immunoreactivity.

From each paraffin block, two sections were prepared—one for H&E staining and one for IHC—each cut at 3-micrometer thickness. IHC sections were mounted on positively charged slides and incubated overnight at 37°C.

For deparaffinization, the slides were immersed in xylene for 30 minutes, followed by treatment with absolute and 96%

ethanol for 15 minutes each. The slides were then rehydrated in distilled water.

Antigen retrieval was conducted using 2 ml of Envision Flex Target Retrieval Solution (50x), High pH (DM 828, Lot No:00096056, Dako), diluted in 98 ml of distilled water. The slides were placed into this solution and heated in a Dako pressure chamber at 1.5 atmospheres for 25 minutes, then allowed to cool at room temperature for 20 minutes. After retrieval, slide borders were outlined with a hydrophobic pen, followed by phosphate-buffered saline (PBS) rinses. The slides were treated with 3% hydrogen peroxide for 20 minutes to block endogenous peroxidase activity. Non-specific binding was prevented using a 15-minute protein block (ScyTek Super Block, Logan, Utah, USA; Ref No: AAA125, Lot No: 21292).

The primary anti-CD1a antibody (Clone EP3622, Cell Marque) was diluted 1:50 and applied to the sections, followed by overnight incubation at 37°C. The following day, slides were rinsed with PBS and incubated with a biotinylated secondary antibody (ScyTek SensiTek Anti-Polyvalent, Biotinylated Antibody; Ref No: ABF125, Lot No: 21020) for 40 minutes. After another PBS wash, streptavidin-conjugated horseradish peroxidase (ScyTek SensiTek HRB; Ref No: ABG125, Lot No: 21026) was applied for 25 minutes. Slides were then stained with aminoethyl carbazole (AEC) chromogen solution (ScyTek Bulk Pack AEC Chromogen/Substrate System; Lot No: 15923), prepared by mixing 20 ml of AEC chromogen with 1 ml of AEC substrate and applied for 15 minutes.

After staining, slides were rinsed in distilled water, counterstained with Mayer's hematoxylin for 5 minutes, rinsed again, and neutralized in ammonia water. Mounting was completed using an aqueous-based mounting medium (Vision Mount, Thermo Scientific/Lab Vision; TA-060-UG, 60 ml).

Evaluation of CD1a Staining

CD1a-positive DCs were identified based on both immunoreactivity and characteristic morphology: intermediate-sized cells with irregular nuclear contours, delicate cytoplasm, and occasional dendritic extensions. In excisional specimens, CD1a-positive cells were manually counted in 50 consecutive HPFs selected from viable tumor regions, excluding necrotic or stromal areas. In small-volume core biopsies, all viable tumor areas were included in the count due to limited tissue availability. This approach was consistent with methods reported in previous studies such as Coventry et al.[6] and Hillenbrand et al.[7] To enhance consistency and reduce observer bias, all cases were jointly evaluated by two pathologists who reached consensus on all immunohistochemical findings. Although formal interobserver variability analysis was not conducted, this collaborative approach helped ensure the reproducibility of results.

The distribution pattern (homogeneous vs. heterogeneous), relationship of DCs with tumor cells, and presence of DCs in the surrounding non-tumoral tissue were also examined.

Statistical Analysis

Descriptive data and immunohistochemical findings were analyzed using SPSS version 16. Relationships between variables were evaluated using Pearson's Chi-square and Fisher's exact tests. OS and DFS analyses were conducted using the Kaplan–Meier method. A p-value <0.05 was considered statistically significant.

RESULTS

All patients included in the study were female, with a median age of 52 years (range: 30–80 years). The majority (43 cases; 89.5%) were diagnosed with invasive ductal carcinoma, while 2 cases (4.16%) had invasive lobular carcinoma, and 3 cases (6.25%) had mixed-type carcinoma with both ductal and lobular components.

All patients had received neoadjuvant chemotherapy due to stage III disease.

The tumors were either strongly HER2-positive (+++) by IHC or moderately positive (++) with HER2 overexpression confirmed by SISH. All patients received trastuzumab, a tyrosine kinase inhibitor, as part of their neoadjuvant regimen. Demographic data and prognostic indicators for patients are summarised in **Table 1**.

Table 1. Prognostic features

Feature	Value
Gender (n=48)	All female
Age (n=48)	Median 52 years (range 30–80 years)
Tumor type (n=48)	Invasive ductal: 43 (89.5%) Invasive lobular: 2 (4.16%) Mixed type: 3 (6.25%)
Tumor size (n=44)	Mean: 3.85 cm Range: 0–18 cm >2 cm: 26 (59.1%) ≤2 cm: 18 (40.9%)
Axillary lymph node status (n=48)	Positive: 26 (54.2%) Negative: 22 (45.8%)
Nuclear grade (n=48)	Grade 3: 37 (77.1%) Grade 2: 11 (22.9%)

CD1a-positive dendritic cell counts in 50 HPFs ranged from 0 to 150, with a median of 7 and a mean of 17.3 cells. No staining was observed in 16 cases (33.3%), while variable levels of CD1a-positive DCs were detected in 32 cases (66.7%). In 15 of these (46.8%), DCs were evenly distributed throughout the tumor tissue; in 17 cases (53.2%), staining was patchy and clustered. In all cases with CD1a positivity, DCs were intermingled with tumor cells, and no DCs were observed in the adjacent non-tumoral tissue (**Figure**). Notably, five cases with CD1a positivity also showed prominent peritumoral lymphocytic infiltration. Data related to CD1a staining are summarised in **Table 2**.

Any level of CD1a positivity was considered when evaluating the association between CD1a expression and survival/prognostic factors. No statistically significant relationship was found between CD1a positivity and age, tumor size, lymph node involvement, or recurrence. No correlation was identified between CD1a positivity/negativity and the neoadjuvant status of the tumor. However, a significant

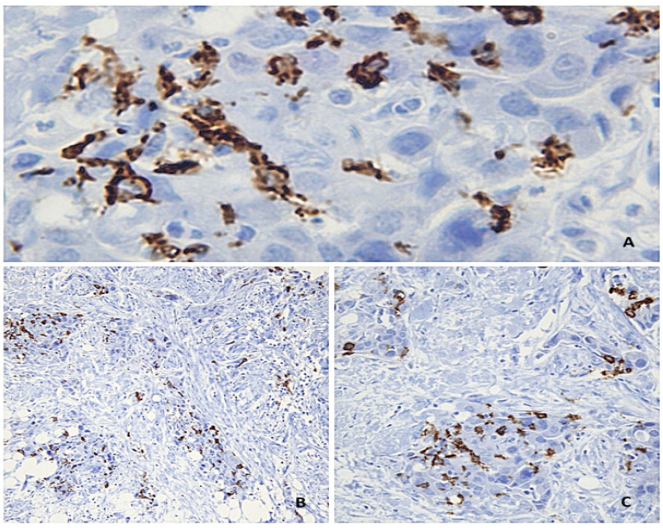


Figure. Immunohistochemical detection of intratumoral CD1a-positive dendritic cells in HER2-positive breast carcinoma. **A.** High-power field (×1000): Numerous CD1a-positive dendritic cells with intense membranous and cytoplasmic staining dispersed among tumor cells. **B.** Low-power field (×200): Diffuse distribution of CD1a-positive cells within tumor clusters. **C.** Intermediate magnification (×400): Focal accumulation of CD1a-positive dendritic cells within viable tumor areas.

Table 2. CD1a-positive dendritic cell distribution

Parameter	Value
CD1a-positive dendritic cell count (in 50 HPFs)	0 to 150
Median count	7
Mean count	17.3 cells
Cases with no staining	16 cases (33.3%)
Cases with CD1a-positive staining	32 cases (66.7%)

association was observed between nuclear grade and CD1a positivity.

CD1a positivity was observed in 36.4% of tumors with nuclear grade 2 and 75.7% of those with nuclear grade 3, indicating a significant correlation between higher nuclear grade and CD1a expression (p=0.015). When patients were grouped according to neoadjuvant therapy status, CD1a positivity was observed in 17 out of 24 patients who did not receive neoadjuvant therapy, and in 15 out of 24 patients who did (p=0.76; Chi-square test), indicating no statistically significant difference between the groups.

OS ranged from 6 to 108 months, with a mean of 55 months. DFS ranged from 1 to 108 months, with a mean of 48 months.

The five-year OS rate was 58% in CD1a-positive cases and 48% in CD1a-negative cases. Although survival appeared better in the CD1a-positive group, the difference was not statistically significant (p=0.73).

Using a cutoff of the median dendritic cell count,[7] the five-year OS was 57% in cases with high CD1a+ cell counts, compared to 52% in those with low or negative counts (p=0.89). ROC analysis was conducted to assess the ability of CD1a-positive cell count to predict DFS. The optimal cut-off determined by the Youden index was 16 cells; however, the AUC was 0.493, indicating no significant predictive value.

Additionally, multiple comparison correction was not applied due to the exploratory nature of the study and limited sample size, which may increase the risk of type I error.

Five-year DFS was 49% in CD1a-positive cases and 36% in CD1a-negative cases (p=0.9). When the threshold was set at 7 (median value) CD1a-positive cells, DFS was 49% in high-expression cases and 41% in low-expression cases. Although there was a proportional difference, it was not statistically significant.

DISCUSSION

This study analyzed 48 non-early-stage cases with HER2 positivity demonstrated by IHC or in situ hybridization. In this regard, the cases included in the study represent a specific subgroup among all breast carcinomas. The primary aim of our study was to evaluate the presence of DCs infiltrating the tumor and, thus, the impact of the immune response against the tumor based on the CD1a molecule in this particular subgroup.

The presence and density of CD1a-positive immature DCs in breast carcinoma tissues vary across studies. Bell et al.[13] applied CD1a and Langerin to 32 breast carcinoma tissues and detected the presence of immature DCs in all cases. DCs ranged from 1 to 48 per 10 HPF, with a median of 7 and a mean of 12. Coventry et al.[6] investigated CD1a (+) DC density in 30 frozen invasive ductal carcinoma tissues and found DCs in half of the samples (0.00–6.05/HPF, 0.00–13.75/mm^2; mean 2.49/mm^2). In another study, Coventry et al.[18] found CD1a (+) DCs in 50% breast cancer tissues, with a mean of 5.5 cells/mm^2. Hillenbrand et al.[7] demonstrated CD1a(+) DCs in 84% of 52 invasive ductal carcinoma tissues. In our study, 32 cases (66.7%) showed variable densities of CD1a (+) DCs. The median number was 7, and the mean was 17.3, consistent with the median reported by Bell et al.,[13] though our mean was higher.

Studies have also examined the distribution and localization of immature DCs and reached similar conclusions. Bell et al.[13] showed that CD1a (+) immature DCs were closely associated with tumor cells, while CD83 (+) mature DCs were primarily found in peritumoral areas near lymphocytic clusters. Coventry et al.[18] observed CD1a (+) DCs loosely clustered in the tumor stroma and concentrated around ductal formations in well-differentiated tumors. Treilleux et al.[19] found close interaction between CD1a (+) and Langerin (+) cells and tumor cells in one-third of 152 patients. Hillenbrand et al.[7] reported higher CD1a (+) DCs in tumor areas than in non-tumor tissues. Similarly, our study found CD1a (+) DCs close to tumor cells. In contrast to the findings of Coventry et al.,[6] no prominent dendritic cell infiltration was observed in the tumor stroma, nor were CD1a (+) DCs present in the peritumoral tissues. The presence of specific mediators may explain this close relationship between DCs and tumor cells. Kradin et al.[20] demonstrated that granulocyte-macrophage colony-stimulating factor (GM-CSF) and IL-2 promote the migration of DCs into tumor areas. Notably, Szpor et al.[21] reported that intratumoral localization of CD1a+ DCs was significantly associated with improved progression-free survival in breast cancer patients, while peritumoral CD1a+

cells lacked such prognostic association. These findings support the potential clinical relevance of dendritic cell localization and align with our observation that CD1a+ cells were predominantly found within the tumor.

The impact of CD1a (+) DCs on overall and DFS has been associated with better prognosis in many cancer types. However, although a proportional association has been observed in breast carcinomas, statistical significance has not been achieved. Coventry et al.[6] studied 48 cases (42 invasive and 6 in situ ductal carcinomas) and found a mortality rate of 32% in cases with low CD1a (+) density, compared to 18% in cases with high density. Nonetheless, this difference was not statistically significant (p=0.331). Treilleux et al.[19] found no significant correlation between CD1a density and overall or DFS in 152 cases of non-metastatic primary breast carcinomas. However, they reported that CD208/DC-LAMP-positive mature DCs were strongly correlated with CD3 (+) T cell infiltration, tumor grade, and lymph node status.

In contrast, the number of immature DCs has been associated with favorable prognosis in many other cancer types. For instance, Furukawa et al.[8] in lung adenocarcinomas, Ambe et al.[22] in colorectal carcinomas, Goldman et al.[12] in tongue carcinomas, and Tsujitani et al.[9] in gastric carcinomas demonstrated this relationship. Eisenthal et al.[23] found that high CD1a (+) DC density reduced recurrence in ovarian carcinomas. Lewko et al.,[24] Lespagnard et al.,[25] and Iwamoto et al.[26] also investigated the prognostic role of immature DCs in breast cancer but found no significant association. However, they noted that CD83(+) mature DCs were related to clinical outcomes.[26]

Our study's 5-year OS rate was 58% in CD1a-positive cases and 48% in CD1a-negative cases. Although this difference was not statistically significant (p=0.73), a proportional trend favoring CD1a positivity was observed. Similar results were obtained using a cutoff of 7 CD1a (+) cells, where higher counts were associated with better outcomes.

Although neoadjuvant therapy has the potential to alter the tumor immune microenvironment, our analysis showed no statistically significant difference in CD1a expression between patients who received neoadjuvant treatment and those who did not (p=0.76). This suggests that neoadjuvant therapy may not markedly affect the presence of CD1a-positive DCs in HER2-positive breast carcinoma tissues, at least within the limitations of our sample size.

Regarding DFS, although not statistically significant (p=0.9), CD1a-positive tumors again showed better 5-year DFS (49%) compared to CD1a-negative tumors (36%). A similar trend was seen with higher dendritic cell counts using the same cutoff.

While many studies suggest that high densities of CD83 (+) mature DCs positively affect survival, the presence of CD1a (+) immature DCs does not show statistically significant survival benefits in breast cancer. This may suggest a possible defect in the maturation of immature DCs in breast carcinomas. Some mediators secreted by tumor cells and the tumor microenvironment have been implicated in this dysfunction. Kradin et al.[20] suggested that IL-10 secreted by tumor cells may inactivate CD1a (+) immature DCs. Gabrilovich et al.[27] also reported VEGF-mediated DC maturation and function suppression.

Studies also investigate the relationship between the presence and density of CD1a (+) DCs and prognostic factors. Coventry et al.[6] found no significant associations with tumor size, grade, lymph node involvement, metastasis, or lymphovascular invasion. Hillenbrand et al.[7] also found no relationship between tumor grade and CD1a (+) cell count. In contrast, our study revealed a statistically significant association between higher nuclear grade and CD1a positivity (p=0.01), which differs from other published reports. No significant association was found with tumor size or axillary lymph node metastasis, consistent with Coventry et al.[6]

Limitations

One of the main limitations of our study is the inclusion of both pre-treatment biopsy specimens and post-neoadjuvant resection materials within the same cohort. Since all patients received trastuzumab-based neoadjuvant therapy, the potential effect of this treatment on CD1a expression remains unclear. This heterogeneity may have influenced the consistency of immunohistochemical findings and limited the comparability of CD1a positivity across subgroups.

While previous studies (e.g., Coventry et al.,[6] Hillenbrand et al.[7]) have assessed CD1a-positive DCs based on quantitative measures such as the number of positive cells per high-power field or per mm^2, an objective semi-quantitative scoring system has not been consistently adopted in the literature. We acknowledge this limitation and have recommended the development of standardized scoring systems in future studies. The absence of a universally accepted cut-off value for CD1a positivity in breast carcinoma posed a significant challenge in our analysis.

Assessing the entire tumor area in small biopsy specimens, while necessary due to limited tissue volume, may have introduced sampling variability when compared to larger excisional specimens. This potential bias is acknowledged as a limitation of the study.

The sample size was limited due to the highly specific nature of the study population. All cases were HER2-positive Stage III breast carcinoma, confirmed by IHC (and SISH when required), treated with trastuzumab-based therapy, and had available archival tissue blocks suitable for immunohistochemical analysis, along with complete 5-year clinical and radiological follow-up data accessible in our institutional records. Although extending the study period could have increased the number of eligible cases, older archival samples often pose challenges in immunohistochemical reliability due to tissue degradation or antigen loss. Therefore, prioritizing both technical feasibility and clinical consistency, we deliberately kept the sample size limited.

This study's retrospective and single-center design, along with variability in follow-up duration among patients, represent potential limitations that may influence the interpretation and generalizability of the findings.

Moreover, although trastuzumab is known to modulate the immune response through mechanisms such as antibody-dependent cellular cytotoxicity, no studies to date have specifically examined its direct effect on CD1a-positive DCs in breast cancer tissues. This lack of evidence limits the interpretation of CD1a expression levels in post-treatment specimens and represents an important area for future research.

CONCLUSION

The cases in our study were all HER2-positive. Considering that HER2-positive tumors account for only 25–30% of breast carcinomas, our study focused on a particular and homogeneous patient group. A lower survival rate was expected because HER2 positivity is a known poor prognostic factor, and our cohort consisted of non-early-stage cases. Therefore, while a proportional relationship between CD1a positivity and survival was observed in HER2-positive cases, statistical significance was not achieved. Future studies should include larger patient populations or more heterogeneous cohorts in terms of prognostic features to obtain more conclusive results. In our study, CD1a positivity was significantly associated only with higher nuclear grade (p=0.015). Thus, based on the data obtained, this specific subgroup warrants further investigation in a larger cohort.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study protocol has been approved by the İstanbul University Clinical Researches Ethics Committee (Date: 04.01.2013, Decision No: 2012/1504-1231).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

This study was supported by İstanbul University Scientific Researches Foundation Grant No: 2013-30198.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

Acknowledgement

The data in this article are taken from the author's thesis titled "Investigation of HER3, CD1A, and Ki-67 expression in HER2-positive breast cancer tissues using immunohistochemical methods, determining the relationship of the results with prognostic parameters." This article is dedicated to the memory of my esteemed thesis advisor, Rıdvan İlhan.

REFERENCES

1. Cella M, Sallusto F, Lanzavecchia A. Origin, maturation and antigen presenting function of dendritic cells. *Curr Opin Immunol.* 1997;9(1):10-16. doi:10.1016/s0952-7915(97)80153-7
2. Söker S. Dendritik hücreler. *Dicle Med J.* 2005;32(3):152-160.
3. Dalod M, Chelbi R, Malissen B, Lawrence T. Dendritic cell maturation: functional specialization through signaling specificity and transcriptional programming. *EMBO J.* 2014;33(10):1104-1116. doi:10.1002/embj.201488027
4. Neutra MR, Frey A, Kraehenbuhl JP. Epithelial m cells: gateways for mucosal infection and immunization. *Cell.* 1996;86(3):345-348. doi:10.1016/s0092-8674(00)80106-3
5. Reid DC. Dendritic cells and immunotherapy for malignant disease. *Br J Haematol.* 2001;112(4):874-887. doi:10.1046/j.1365-2141.2001.02626.x
6. Coventry BJ, Morton J. Cd1a-positive infiltrating-dendritic cell density and 5-year survival from human breast cancer. *Br J Cancer.* 2003;89(3):533-538. doi:10.1038/sj.bjc.6601114
7. Hillenbrand EE, Neville AM, Coventry BJ. Immunohistochemical localization of cd1a-positive putative dendritic cells in human breast tumours. *Br J Cancer.* 1999;79(5-6):940-944. doi:10.1038/sj.bjc.6690150
8. Furukawa T, Watanabe S, Kodama T, Sato Y, Shimosato Y, Suemasu K. T-zone histiocytes in adenocarcinoma of the lung in relation to postoperative prognosis. *Cancer.* 1985;56(11):2651-2656.doi:10.1002/1097-0142(19851201)56:11<2651::aid-cncr2820561120>3.0.co;2-e
9. Tsujitani S, Furukawa T, Tamada R, Okamura T, Yasumoto K, Sugimachi K. Langerhans cells and prognosis in patients with gastric carcinoma. *Cancer.* 1987;59(3):501-505. doi:10.1002/1097-0142(19870201)59:3<501::aid-cncr2820590325>3.0.co;2-h
10. Gallo O, Bianchi S, Giannini A, Gallina E, Libonati GA, Fini-Storchi O. Correlations between histopathological and biological findings in nasopharyngeal carcinoma and its prognostic significance. *Laryngoscope.* 1991;101(5):487-493. doi:10.1288/00005537-199105000-00008
11. Gallo O, Libonati GA, Gallina E, et al. Langerhans cells related to prognosis in patients with laryngeal carcinoma. *Arch Otolaryngol Head Neck Surg.* 1991;117(9):1007-1010. doi:10.1001/archotol.1991.01870210079015
12. Goldman SA, Baker E, Weyant RJ, Clarke MR, Myers JN, Lotze MT. Peritumoral cd1a-positive dendritic cells are associated with improved survival in patients with tongue carcinoma. *Arch Otolaryngol Head Neck Surg.* 1998;124(6):641-646. doi:10.1001/archotol.124.6.641
13. Bell D, Chomarat P, Broyles D, et al. In breast carcinoma tissue, immature dendritic cells reside within the tumor, whereas mature dendritic cells are located in peritumoral areas. *J Exp Med.* 1999;190(10):1417-1426. doi:10.1084/jem.190.10.1417
14. Strunk D, Egger C, Leitner G, Hanau D, Stingl G. A skin homing molecule defines the langerhans cell progenitor in human peripheral blood. *J Exp Med.* 1997;185(6):1131-1136. doi:10.1084/jem.185.6.1131
15. Richters CD, Hoekstra MJ, van Baare J, Du Pont JS, Hoefsmit EC, Kamperdijk EW. Isolation and characterization of migratory human skin dendritic cells. *Clin Exp Immunol.* 1994;98(2):330-336. doi:10.1111/j.1365-2249.1994.tb06146.x
16. Hicks DG, Kulkarni S. Her2+ breast cancer: review of biologic relevance and optimal use of diagnostic tools. *Am J Clin Pathol.* 2008;129(2):263-273. doi:10.1309/99ae032r9fm8wnd1
17. Bloom HJ, Richardson WW. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. *Br J Cancer.* 1957;11(3):359-377. doi:10.1038/bjc.1957.43
18. Coventry BJ, Lee PL, Gibbs D, Hart DN. Dendritic cell density and activation status in human breast cancer--cd1a, cmrf-44, cmrf-56 and cd-83 expression. *Br J Cancer.* 2002;86(4):546-551. doi:10.1038/sj.bjc.6600132
19. Treilleux I, Blay JY, Bendriss-Vermare N, et al. Dendritic cell infiltration and prognosis of early stage breast cancer. *Clin Cancer Res.* 2004;10(22):7466-7474. doi:10.1158/1078-0432.ccr-04-0684
20. Kradin RL, Xia W, Pike M, Byers HR, Pinto C. Interleukin-2 promotes the motility of dendritic cells and their accumulation in lung and skin. *Pathobiology.* 1996;64(4):180-186. doi:10.1159/000164033
21. Szpor J, Streb J, Glajcar A, et al. Dendritic cells are associated with prognosis and survival in breast cancer. *Diagnostics (Basel).* 2021;11(4):702. doi:10.3390/diagnostics11040702

22. Ambe K, Mori M, Enjoji M. S-100 protein-positive dendritic cells in colorectal adenocarcinomas. Distribution and relation to the clinical prognosis. *Cancer.* 1989;63(3):496-503. doi:10.1002/1097-0142(19890201)63:3<496::aid-cncr2820630318>3.0.co;2-k
23. Eisenthal A, Polyvkin N, Bramante-Schreiber L, Misonznik F, Hassner A, Lifschitz-Mercer B. Expression of dendritic cells in ovarian tumors correlates with clinical outcome in patients with ovarian cancer. *Hum Pathol.* 2001;32(8):803-807. doi:10.1053/hupa.2001.26455
24. Lewko B, Zółtowska A, Stepiński J, Roszkiewicz A, Moszkowska G. Dendritic and cancer cells in the breast tumors--an immunohistochemical study: short communication. *Med Sci Monit.* 2000;6(5):892-895.
25. Lespagnard L, Gancberg D, Rouas G, et al. Tumor-infiltrating dendritic cells in adenocarcinomas of the breast: a study of 143 neoplasms with a correlation to usual prognostic factors and to clinical outcome. *Int J Cancer.* 1999;84(3):309-314. doi:10.1002/(sici)1097-0215(19990621)84:3<309::aid-ijc19>3.0.co;2-3
26. Iwamoto M, Shinohara H, Miyamoto A, et al. Prognostic value of tumor-infiltrating dendritic cells expressing cd83 in human breast carcinomas. *Int J Cancer.* 2003;104(1):92-97. doi:10.1002/ijc.10915
27. Gabrilovich DI, Chen HL, Girgis KR, et al. Production of vascular endothelial growth factor by human tumors inhibits the functional maturation of dendritic cells. *Nat Med.* 1996;2(10):1096-1103. doi:10.1038/nm1096-1096

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1741140

Anatolian Curr Med J. 2025;7(5):630-634

Multidrug-resistant bacteria in hospitalized pediatric patients with cystic fibrosis: a retrospective cross-sectional study

Serkan Atıcı[1], Emine Atağ[2], Bülent Karadağ[3], Ahmet Soysal[4]

[1]*Division of Pediatric Infectious Diseases, Department of Pediatrics, Faculty of Medicine, Okan University, İstanbul, Turkiye*
[2]*Division of Pediatric Pulmonology, Department of Pediatrics, Faculty of Medicine, Maltepe University, İstanbul, Turkiye*
[3]*Division of Pediatric Pulmonology, Department of Pediatrics, Faculty of Medicine, Marmara University, İstanbul, Turkiye*
[4]*Division of Pediatric Infectious Diseases, Department of Pediatrics, Memorial Ataşehir Hospital, İstanbul, Turkiye*

Cite this article as: Atıcı S, Atağ E, Karadağ B, Soysal A. Multidrug-resistant bacteria in hospitalized pediatric patients with cystic fibrosis: a retrospective cross-sectional study. *Anatolian Curr Med J.* 2025;7(5):630-634.

Received: 13.07.2025 ◆ **Accepted**: 01.09.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: The emergence of multidrug-resistant (MDR) bacterial pathogens has become a critical clinical challenge in cystic fibrosis (CF) patients. This study aimed to characterize MDR bacterial isolates in hospitalized CF patients experiencing pulmonary exacerbations.

Methods: We defined MDR as resistance to three or more antimicrobial classes. The study included 45 acute exacerbation episodes with confirmed MDR bacterial growth among 32 CF pediatric patients hospitalized at Marmara University Pendik Training and Research Hospital.

Results: Among 32 CF patients (mean age: 114.7±62.1 months; 68.8% female), we analyzed 45 exacerbation episodes. The most common clinical manifestations were increased sputum production (73.3%), cough (60%), and fever (31%). *Pseudomonas aeruginosa* colonization was previously documented in 55.6% of cases. Respiratory cultures revealed MDR bacteria in all patients, with *P. aeruginosa* (53.4%), *Staphylococcus aureus* (17.8%), and *Acinetobacter baumannii* (4.4%) being the predominant isolates. Combination antibiotic therapy was required in 82% of cases (37/45), while only 18% (8/45) received monotherapy. The mean hospitalization duration was 17.3±8.4 days, with an overall mortality rate of 6.2% (n=2).

Conclusion: MDR infections in CF patients demand a multifaceted approach. Rotating antibiotic regimens and avoiding carbapenem overuse may reduce resistance. Tailored antibiotic regimens and antimicrobial stewardship programme, robust infection controls such as enhanced screening for MDR colonization and isolation protocols are needed to curb transmission, and sustained research into emerging pathogens. Our data reinforce the urgent need for CF-specific antimicrobial protocols in Turkiye and similar countries.

Keywords: Cystic fibrosis, multidrug-resistant, hospitalized patients

INTRODUCTION

Cystic fibrosis (CF) is a highly prevalent autosomal recessive and life-limiting disorder among Caucasian populations. Dysfunction of the CF transmembrane conductance regulator (CFTR) in airway epithelial cells disrupts sodium chloride and water transport, impairing mucociliary clearance and leading to viscous mucus accumulation.[1] This pathophysiology facilitates opportunistic pathogen colonization and chronic infection, resulting in recurrent respiratory infections and progressive lung function decline. While advancements such as CFTR modulators, respiratory physiotherapy, airway clearance techniques, and optimized nutritional support have reduced morbidity and improved survival rates, in pediatric CF patients, infections, particularly those caused by multidrug-resistant (MDR) microorganisms, remain a major cause of morbidity and mortality. Prompt and effective treatment of infective exacerbations is critical to mitigate accelerated lung function decline.[2,3]

Antimicrobial resistance has emerged as one of the top global public health threats of the 21st century, with the World Health Organization (WHO) classifying it as a "silent pandemic" that could cause 10 million deaths annually by 2050 if unchecked.[4] Among high-risk populations, CF patients are particularly vulnerable due to chronic airway infections, frequent antibiotic exposure, and prolonged hospitalizations, key drivers of resistance development. The term "MDR" lacks a universal definition but is most commonly applied to pathogens resistant to ≥3 antimicrobial classes.[5] In CF, MDR pathogens pose unique challenges. *Pseudomonas aeruginosa* up to 30% of CF isolates are MDR, with resistance to carbapenems, fluoroquinolones, and aminoglycosides.[6]

Corresponding Author: Serkan Atıcı, drserkanatici@gmail.com

Rising methicillin-resistant *Staphylococcus aureus* (MRSA), *Acinetobacter baumannii* and *Stenotrophomonas maltophilia* rates correlate with worse lung function and reduce antibiotic treatment options. Broader spectrum and repeated antibiotic use may predispose to disruption of the microbiota, fungal overgrowth (e.g., *Candida*, and *Aspergillus*) and further resistance. MDR infections could increase mortality risk by 2- to 3-fold in CF patients.[7] While MDR trends in CF are well-documented in United States (U.S.) and Europe, data from Turkiye and similar middle-income countries remain limited. Our study aimed to contribute to filling this gap by analyzing the characteristics of MDR pathogens in CF.

METHODS

This retrospective study was conducted in accordance with recognized ethical standards, including the principles of the Declaration of Helsinki, and received approval from the Memorial Ataşehir Hospital Non-interventional Clinical Researches Ethics Committee (Date: 16.06.2025, Decision No: 07.2025).

This retrospective cohort study was conducted at Marmara University Pendik Training and Research Hospital, a tertiary care center in İstanbul, Turkiye, between January 2013–December 2015. CF patients who met modified Fuchs criteria for pulmonary exacerbation, such as increased cough/sputum production, ≥10% decrease in FEV1 (if present), new radiographic findings, and were hospitalized were included in the study.[8] Study was performed on patients with microbiologically confirmed MDR pathogen growth in respiratory samples (sputum, nasopharyngeal aspirate, and bronchoalveolar lavage). Patients with incomplete medical records and non-MDR infection episodes were excluded from the study. Each exacerbation episode was considered a separate case; as some patients contributed more than one episode, this was taken into account when interpreting the findings. Demographic data such as age, sex; clinical data such as patient symptoms, previous colonization status, antibiotic treatments, duration of hospitalization; and microbiological data such as culture results and antibiotic susceptibility tests were extracted from the electronic patient system using a standardized form. The following definitions were used in the study;

Prior colonization: Isolation of the same pathogen in the past 6 months. Combination therapy: Concurrent use of ≥2 antibiotics from different classes. MDR: Resistance to ≥3 antimicrobial classes (e.g., β-lactams, aminoglycosides, fluoroquinolones).[5]

The identification of bacterial species was performed using the VITEK2 system (BioMérieux, Marcy l'Etoile, France). Susceptibility to antibiotics was evaluated using the disk diffusion method according to the criteria set by the Clinical and Laboratory Standards Institute (CLSI).[9]

Statistical Analysis

The data were entered into Microsoft Office Excel (Microsoft Corp., USA) and analyzed via the SPSS for Windows (SPSS, Chicago, IL, USA) program. Mean±SD for continuous variables (e.g., age), frequencies (%) for categorical data (e.g., pathogen distribution). Comparative analyses: Chi-square/Fisher's exact tests for associations (e.g., mortality vs. antibiotic regimen). Mean±SD for continuous variables (e.g., age, hospitalization duration), frequencies (%) for categorical data (e.g., pathogen distribution) were used.

RESULTS

A total of 32 CF patients experiencing 45 acute pulmonary exacerbation episodes were included in this study. Each exacerbation episode was considered a separate case; as some patients contributed more than one episode, this was taken into account when interpreting the findings. The cohort exhibited a female predominance, with 22 females (68.8%) and 10 males (31.2%), a distribution consistent with epidemiological trends in CF populations. The mean age of the patients was 114.7±62.1 months (approximately 9.5 years), with an age range of 5 to 216 months. At the time of hospitalization, the most frequently reported clinical symptoms were increased sputum production (73.3%, n=33/45), cough (60%, n=27/45), and fever (31%, n=14/45). The high prevalence of increased sputum production aligns with the known pathophysiology of CF pulmonary exacerbations, where mucus hypersecretion and impaired clearance contribute to bacterial retention and infection.

The study included CF patients with MDR bacterial growth, and microbiological analysis confirmed MDR bacterial isolates in respiratory samples from all 45 exacerbation episodes. *Pseudomonas aeruginosa* was the most prevalent pathogen, isolated in 53.4% (n=24/45) of cases. Notably, 62.5% (n=15/24) of these episodes occurred in patients with pre-existing *P. aeruginosa* colonization, suggesting that chronic colonization is a significant risk factor for MDR development in CF patients. *Staphylococcus aureus* accounted for 17.8% (n=8/45) of isolates, with equal distribution between methicillin-susceptible (MSSA, 8.9%; n=4/45) and methicillin-resistant (MRSA, 8.9%; n=4/45) strains. *Acinetobacter baumannii*, a less common but emerging pathogen in CF, was identified in 4.5% (n=2/45) of cases. In 17.8% (n=8/45) of episodes, more than one pathogens were isolated, with combinations such as *P. aeruginosa*+ MSSA or *P. aeruginosa*+ *Haemophilus influenzae*, complicating treatment decisions (**Figure**).

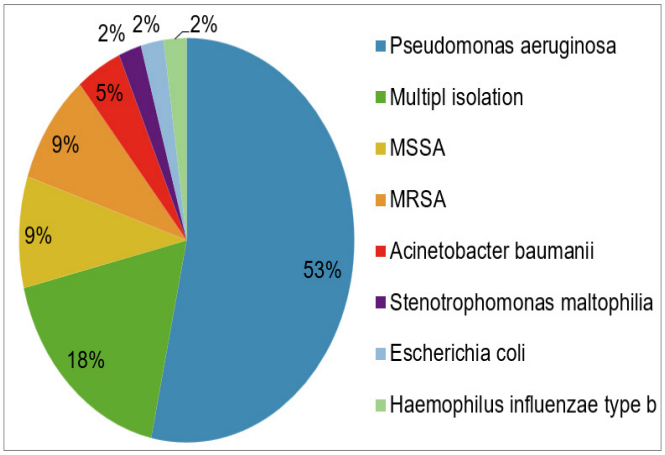


Figure. Distribution of isolates
MSSA: Methicillin sensitive *staphylococcus aureus*, MRSA: Methicillin resistant *staphylococcus aureus*, Multipl isolation including *P. aeruginosa*, MSSA, *Haemophilus influenzae* type b and *Stenotrophomonas maltophilia*

P. aeruginosa isolates exhibited high resistance rates to aminoglycosides (54.2%, n=13/24) and carbapenems (45.8%, n=11/24). However, they demonstrated greater susceptibility to quinolones (e.g., ciprofloxacin and levofloxacin: 29.2%, n=7/24) and remained susceptible to polymyxins (e.g., colistin: 0% resistance). *S. aureus* isolates showed universal susceptibility to vancomycin, but MRSA strains (75%, n=3/4) were resistant to β-lactams, necessitating alternative regimens. Resistance rates were also analyzed according to antibiotic classes regardless of microorganism type. As shown in **Table**, cephalosporins demonstrated the highest resistance rate (66.7%, 30/45 isolates), followed by aminoglycosides (40.0%, 18/45). Carbapenem, quinolone, and sulfonamides resistance was observed in 28.9% (13/45), 22.2% (10/45), and 17.8% (8/45) of isolates respectively. Confidence intervals indicate significant variability in resistance prevalence across antibiotic classes.

Table. Resistance patterns of bacterial isolates to major antibiotic classes

Antibiotic class	Resistant isolates (n=45)	Percentage resistance (95% CI)
Cephalosporins	30	66.7% (51.7-79.0)
Aminoglycosides	18	40.0% (26.8-55.2)
Carbapenems	13	28.9% (17.5-43.9)
Quinolones	10	22.2% (12.3-36.6)
Sulfonamides	8	17.8% (9.4-25.9)

CI: Confidence interval

The management of MDR-related exacerbations required individualized antibiotic regimens, antibiotic treatments were arranged according to the patient's clinical and microbiological findings. Monotherapy was employed in only 18% (n=8/45) of cases, typically for infections caused by pathogens with limited resistance profiles (e.g., MSSA). Combination therapy was the mainstay of treatment, used in 82% (n=37/45) of episodes. The most frequently prescribed combination was ceftazidime+amikacin (22.2%, n=10/45), chosen for its synergistic activity against *P. aeruginosa*. Other regimens included meropenem+colistin (8.9%, n=4/45) and other carbapenem-based combinations (26.7%, n=12/45). Due to concerns about fungal superinfections, empirical antifungal therapy (e.g., fluconazole, voriconazole) was administered in 17.8% (n=8/45) of cases, though confirmatory fungal cultures were not always obtained. The mean hospitalization duration for exacerbations was 17.3±8.4 days, with a range of 10 to 60 days. Mortality was observed in 6.2% (n=2/32) of patients. Both fatal cases exhibited *P. aeruginosa* infections. The first patient received combination therapy with meropenem, amikacin, and ciprofloxacin, while the second was received meropenem-colistin dual therapy.

DISCUSSION

This single-center retrospective study provides critical insights into the burden of MDR bacterial infections in hospitalized CF patients. The study highlights the alarming prevalence of MDR bacterial infections in hospitalized CF patients with pulmonary exacerbations, with *Pseudomonas aeruginosa* (53.4%) as the dominant pathogen. This aligns with CF Foundation Registry data, which reports *P. aeruginosa* colonization in ~50-60% of CF patients.[10] The high rate of prior colonization (55.6% in our cohort) underscores the chronicity of this pathogen in CF airways and its role in driving recurrent exacerbations. The predominance of *P. aeruginosa* in our cohort parallels studies from other regions (e.g., UK CF Registry 2022), but the 17.8% prevalence of *Staphylococcus aureus* (including MRSA) is notably higher than in some European cohorts (~10%).[11,12] This discrepancy may reflect regional differences in antibiotic stewardship or infection control practices. The isolation of *Acinetobacter baumannii* (4.4%), a rare but emerging CF pathogen, warrants vigilance, as its intrinsic resistance to multiple antibiotics complicates treatment, as reported by Rocha et al.[13] Studies from Southern Europe and Asia report *A. baumannii* colonization rates of 3–8% in CF patients, often linked to prolonged hospital stays and ventilator use.[14] Our findings mirror this trend, suggesting regional spread. Unlike *P. aeruginosa*, which adapts to CF airways over years, *A. baumannii* may exploit immune compromise or prior antibiotic exposure.

Notably, 62.5% of *P. aeruginosa* isolates emerged from patients with prior colonization, consistent with studies linking chronic colonization to MDR development.[3] The high resistance rates to aminoglycosides (54.2%) and carbapenems (45.8%) mirror trends reported in Europe and the U.S., likely driven by frequent antibiotic exposure.[4] The high resistance to aminoglycosides (e.g., amikacin, gentamicin) aligns with global reports of *P. aeruginosa* developing efflux pumps (e.g., mexXY-oprM) and ribosomal modifications (rpsL mutations).[1] In CF, chronic aminoglycoside use for suppression therapy may drive this trend.[2] Notably, our cohort's resistance rate (54.2%) exceeds the 30–40% reported in recent U.S. CF Registry data, possibly reflecting regional prescribing practices, lack of antimicrobial stewardship programme, and inadequate infection control measures.[3] However, the 0% resistance to colistin offers a therapeutic silver lining, supporting its use as a last-resort agent.[5] The equal distribution of MSSA and MRSA (8.9% each) contrasts with higher MRSA rates in some Western cohorts, possibly reflecting regional prescribing practices or infection control measures.[6] The absence of vancomycin resistance in *S. aureus* isolates is reassuring, though MRSA's β-lactam resistance necessitates alternative regimens.

The frequent need for combination therapy (82%), particularly ceftazidime-amikacin (22.2%), highlights the complexity of MDR infections. This aligns with CF treatment guidelines advocating synergistic combinations to overcome resistance.[2] However, the 28.9% carbapenem resistance raises concerns about overuse, echoing calls for antimicrobial stewardship in CF centers.[8] The 17.8% empirical antifungal use without confirmed fungal co-infections suggests clinical caution, though unnecessary antifungals may exacerbate resistance or adverse effects.[5] Future studies should assess fungal co-infection rates to guide practice. The mean hospitalization duration (17.3 days) exceeds reports from centers with stricter MDR protocols, suggesting room for improvement in early diagnosis and targeted therapy.[15] Both fatalities involved MDR *P. aeruginosa*, consistent with studies linking this pathogen to poorer outcomes.[16] The first patient

received meropenem-amikacin-ciprofloxacin, while the second received meropenem-colistin, a regimen typically reserved for extensively drug-resistant cases. Their deaths despite aggressive therapy underscore the urgency of novel antimicrobial strategies, such as phage therapy or anti-biofilm agents.[17] Due to the small sample size, no statistically significant association between mortality and treatment type could be demonstrated.

Limitations

This study has several limitations. Its retrospective design may introduce documentation bias; the single-center cohort limits generalizability; and small sample size precludes subgroup analyses (e.g., genotype-specific resistance patterns). Additionally, molecular characterization of resistance mechanisms (e.g., blaKPC, mexAB-oprM) could not be performed. Nevertheless, the detailed microbiological data and real-world treatment patterns provide valuable insights for regional clinicians. Another limitation of our study is the period of data collection (2013–2015). Since then, advances in CF care, particularly the introduction of CFTR modulators and evolving antimicrobial stewardship strategies, may have altered the prevalence and resistance patterns of MDR pathogens. Therefore, our findings may not fully represent the current epidemiology of MDR infections in CF patients, especially in countries where novel therapies have become more accessible. Nevertheless, the study provides valuable baseline information from a period preceding the widespread use of these therapies, which may serve as a benchmark for future comparative studies. Another important limitation is the study included only MDR-positive exacerbations; MDR-negative episodes were not analyzed as a separate group. Consequently, the direct impact of MDR status on patient outcomes could not be assessed. Future prospective studies should address this gap by comparing MDR-positive and MDR-negative exacerbations in order to better quantify the attributable risk.

CONCLUSION

MDR bacterial infections remain a major challenge in hospitalized pediatric CF patients, with *Pseudomonas aeruginosa* as the predominant pathogen. High resistance rates, particularly to aminoglycosides and carbapenems, underscore the urgent need for antimicrobial stewardship and tailored combination therapies. Our findings highlight the importance of early detection, strict infection control, and judicious antibiotic use to curb resistance. Future multicenter studies comparing MDR-positive and MDR-negative exacerbations, and evaluating novel therapies, are needed. These results emphasize the necessity of CF-specific antimicrobial protocols in Turkiye and similar settings.

ETHICAL DECLARATIONS

Ethics Committee Approval

Received approval from the Memorial Ataşehir Hospital Non-interventional Clinical Researches Ethics Committee (Date: 16.06.2025, Decision No: 07.2025).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

Acknowledgement

We would like to thanks Dr. Ibrahim Aminu for his editorial contributions, including the English language evaluation.

REFERENCES

1. Aydemir Ö, Kiper N. Cystic fibrosis disease in the world and Turkey. *J Int Med Sci*. 2007;3(14):1-3.
2. Conway SP, Brownlee KG, Denton M, Peckham DG. Antibiotic treatment of multidrug-resistant organisms in cystic fibrosis. *Am J Respir Med*. 2003;2(4):321-332. doi:10.1007/BF03256660
3. Fainardi V, Neglia C, Muscarà M, et al. Multidrug-resistant bacteria in children and adolescents with cystic fibrosis. *Children (Basel)*. 2022;9(9):1330. doi:10.3390/children9091330
4. World Health Organization. Global antimicrobial resistance and use surveillance system (GLASS) report: 2021. [accessed: May 10, 2025] https://iris.who.int/handle/10665/341666.
5. Magiorakos AP, Srinivasan A, Carey RB, et al. Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin Microbiol Infect*. 2012;18(3):268-281. doi:10.1111/j.1469-0691.2011.03570.x
6. Cromwell EA, Ostrenga JS, Todd JV, et al. Cystic fibrosis prevalence in the United States and participation in the cystic fibrosis foundation patient registry in 2020. *J Cyst Fibros*. 2023;22(3):436-442. doi:10.1016/j.jcf.2023.02.009
7. Wood SJ, Kuzel TM, Shafikhani SH. *Pseudomonas aeruginosa*: infections, animal modeling, and therapeutics. *Cells*. 2023;12(1):199. doi:10.3390/cells12010199
8. Almulhem M, Ward C, Haq I, Gray RD, Brodlie M. Definitions of pulmonary exacerbation in people with cystic fibrosis: a scoping review. *BMJ Open Respir Res*. 2024;11(1):e002456. doi:10.1136/bmjresp-2024-002456
9. Clinical and Laboratory Standards Institute. Performance standards for antimicrobial susceptibility testing. 25th informational supplement. CLSI document M100-S25. Wayne, PA: CLSI 2015.
10. Cystic Fibrosis Foundation. 2023 Patient Registry Annual Data Report [Internet]. Bethesda, MD: Cystic Fibrosis Foundation; 2024 [cited 2025 May 10]. Available from: https://www.cff.org/medical-professionals/patient-registry
11. UK Cystic Fibrosis Registry. 2022 Annual Data Report [Internet]. London: UK Cystic Fibrosis Trust; 2023 [cited 2025 May 10]. Available from: https://www.cysticfibrosis.org.uk/the-work-we-do/uk-cf-registry/reports-and-data
12. European Cystic Fibrosis Society. Patient Registry 2022: Key Findings [Internet]. Copenhagen: ECFS; 2023 [cited 2025 May 10]. Available from: https://www.ecfs.eu/projects/patient-registry

13. Rocha GA, Lima DF, Rodrigues ER, et al. Species distribution, sequence types and antimicrobial resistance of *Acinetobacter* spp. from cystic fibrosis patients. *Epidemiol Infect*. 2018;146(4):524-530. doi:10.1017/S0950268817002849
14. Moubareck CA, Halat DH. Insights into *Acinetobacter baumannii*: a review of microbiological, virulence, and resistance traits in a threatening nosocomial pathogen. *Antibiotics (Basel).* 2020;9(3):119. doi:10.3390/antibiotics9030119
15. Ranganathan SC, Hall GL, Sly PD, Stick SM, Douglas TA; AREST CF. Early lung disease in infants and preschool children with cystic fibrosis: what have we learned, and what should we do about it? *Am J Respir Crit Care Med*. 2017;195(12):1567-1575. doi:10.1164/rccm.201606-1107CI
16. Smyth AR, Bell SC, Bojcin S, et al. European Cystic Fibrosis Society standards of care: best practice guidelines. *J Cyst Fibros*. 2014;13(Suppl 1):S23-42. doi:10.1016/j.jcf.2014.03.010
17. Waters V, Smyth A. Cystic fibrosis microbiology: advances in antimicrobial therapy. J *Cyst Fibros*. 2015;14(5):551-560. doi:10.1016/j.jcf.2015.02.005

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1757798

Anatolian Curr Med J. 2025;7(5):635-641

Long-term regular exercise effect on retinal and choroidal structure: insights from real-life data

Nur Demir[1], Hasine Gözde Dalkılıç[1], Belma Kayhan[1], Nurgül Tükel[2], Zeliha Serindağ[2], Mustafa Kaplan[2]

[1]*Department of Ophthalmology, Sultan 2. Abdulhamid Han Training and Research Hospital, University of Health Sciences, İstanbul, Turkiye*
[2]*Department of Internal Medicine, Sultan 2. Abdulhamid Han Training and Research Hospital, University of Health Sciences, İstanbul, Turkiye*

Cite this article as: Demir N, Dalkılıç HG, Kayhan B, Tükel N, Serindağ Z, Kaplan M. Long-term regular exercise effect on retinal and choroidal structure: insights from real-life data. *Anatolian Curr Med J.* 2025;7(5):635-641.

Received: 04.08.2025 • **Accepted**: 01.09.2025 • **Published**: 15.09.2025

ABSTRACT

Aims: Exercise increases ocular perfusion pressure. The retina and choroid can autoregulate blood flow to a certain extent in response to this increase. Most of the studies have focused on short-term ophthalmologic effects of exercise and showed conflicting results. This study aimed to evaluate the long-term effects of regular exercise on retinal and choroidal structures, as well as the potential contribution of exercise-related systemic changes.

Methods: Participants who had engaged in regular exercise for at least two years were included in the study group, while those who had not exercised were assigned to the control group. Clinical assessments included blood pressure, heart rate, body-mass index, HbA1c, complete blood count, lipid profile, thyroid-stimulating hormone, fT3, fT4, and C-reactive protein. Retinal and choroidal thicknesses in the right eye were measured using optical coherence tomography.

Results: The trained group consisted of 36 participants and the untrained group included 35 participants. The mean duration of regular exercise was 4.29±2.5 years. Choroidal thickness was greater at all measured points in the trained group compared with controls, with significant differences at nasal ($p<0.05$) and subfoveal ($p<0.01$) measurements. Red blood cell count, hemoglobin, hematocrit, mean corpuscular volume, mean corpuscular hemoglobin, and eosinophil counts were also significantly higher in the trained group.

Conclusion: This study is, to our knowledge, the first to investigate the effects of long-term, consistent physical activity by non-professionals in real-life settings on retinal and choroidal structures. Choroidal autoregulation normally maintains stable blood flow even when ocular perfusion pressure rises during exercise. However, the observed long-term choroidal thickening suggests that prolonged physical activity may exceed this regulatory capacity.

Keywords: Cardiovascular risk factors, choroid, exercise, hematological parameters, real-life conditions, retina

INTRODUCTION

During physical activity, ocular blood flow increases in parallel with rises in blood pressure (BP) and heart rate (HR).[1] It has been shown in many studies that autoregulatory mechanisms prevent the increase of ocular blood flow during exercise.[2] Retinal vessels have no innervation and retinal blood flow is regulated with oxygen, carbon dioxide and local metabolic factors.[1] The typical retinal response to exercise is vasoconstriction.[1,3]

The choroid is a highly vascularized layer of the eye.[4] Its vascular bed receives both sympathetic and parasympathetic innervation, with sympathetic input and ocular perfusion pressure (OPP) serving as the main regulatory factors.[4] Moreover, abnormalities in blood count, glucose metabolism, BP, and lipid profile can influence choroidal structure.[5] Therefore, the choroid is likely to be affected by cardiovascular risk factors.[5] Variations in hemogram parameters may also alter choroidal blood flow (CBF) and thickness.[6]

Regular exercise exerts systemic effects on the arterial vasculature.[7,8] It reduces total peripheral resistance and cardiac afterload; however, these are not the only mechanisms through which exercise lowers BP.[7] Regular training also decreases sympathetic activity, prevents arterial stiffness, and contributes to BP reduction by mitigating inflammation.[7] In addition, exercise improves traditional cardiovascular risk factors, including insulin resistance, hypertension, dyslipidemia, and obesity.[8,9]

There are only a few studies evaluating the long-term effects of exercise on the retina and choroid. The aim of this study is to evaluate the impact of regular exercise on the retinochoroidal structure of healthy people under real-life conditions. In

Corresponding Author: Nur Demir, nurd9920@gmail.com

addition, the study was designed to assess systemic parameters that may influence the retina and choroid in conjunction with exercise.

METHODS

This is a cross-sectional study that enrolled voluntary participants of the same ethnicity aged 18 to 50 years. This study was approved by the University of Health Sciences Hamidiye Clinical Researches Ethics Committee (Date: 23.11.2023, Decision No: 2023.23.11-63) and conducted in accordance with the Declaration of Helsinki. Informed consent was obtained from all participants. Best corrected visual acuity, intraocular pressure (IOP), biomicroscopic anterior segment, and pupil-dilated posterior segment examinations were performed.

Participants who had engaged in regular exercise for ≥2 years and accumulated >1500 metabolic equivalent of task (MET) per week were included. Physical activity was assessed using the Global Physical Activity Questionnaire (GPAQ). One MET denotes resting energy expenditure (≈3.5 ml $O_2 \cdot kg^{-1} \cdot min^{-1}$ ≈1 $kcal \cdot kg^{-1} \cdot h^{-1}$). Moderate-intensity activities (e.g., recreational cycling, dancing at a moderate pace) were assigned a value of 4 METs, and vigorous-intensity activities (e.g., aerobics, basketball, high-intensity fitness training) 8 METs. The weekly total MET value was calculated as the sum across all reported activities: exercise constant (4 or 8)×exercise duration (minutes)×exercise frequency (days per week).

Biochemical and hematological parameters were measured using standardized kits and automated analyzers. After an overnight fast of at least eight hours, peripheral blood samples were obtained from all participants. The analyses included HbA1c, complete blood count (CBC), lipid profile, thyroid-stimulating hormone (TSH), fT3, fT4, and C-reactive protein (CRP). The lipid profile comprised total cholesterol, HDL-cholesterol, LDL-cholesterol, and triglycerides. CBC analysis was performed with the Mindray BC-6800Plus hematology analyzer using SF Cube technology. The lipid profile was assessed with the Roche Lipid Panel on the cobas c 501 analyzer using an enzymatic colorimetric method. HbA1c was measured with the Cobas b 101 system, which employs an immunoassay method. Thyroid function tests were measured using the Elecsys TSH assay on the cobas e 601 analyzer based on the electrochemiluminescence immunoassay technique.

HR, BP, height, and weight were measured for all participants, and body-mass index (BMI) was calculated as weight in kilograms divided by the square of height in meters. The standard protocol for office BP monitoring included a 5-minute rest period before measurement, abstinence from smoking, alcohol, or caffeine for at least 30 minutes, and avoidance of talking during and between measurements. BP measurements were obtained using automated oscillometric sphygmomanometers. Systolic blood pressure (SBP), diastolic blood pressure (DBP), and HR were recorded from the upper arm using the Omron M6 Comfort HEM-7360-E device.

The exclusion criteria included refractive errors greater than 3 diopters of myopia, hyperopia, and astigmatism, glaucoma, ocular surgery, retinal diseases, diabetes, respiratory, thyroid, and cardiovascular diseases, hypertension, vascular diseases, nephropathy, uveitis, cigarette smoking, alcohol intake, underweight BMI (<18.5 kg/m^2); and overweight BMI (>25 kg/m^2).

The right eyes of the patients were evaluated. Optical coherence tomography (OCT) measured total retinal thickness and retinal layer thicknesses, including the nerve fiber layer (NFL), inner plexiform layer (IPL), ganglion cell layer (GCL), inner nuclear layer (INL), outer plexiform layer (OPL), outer nuclear layer (ONL), and retinal pigment epithelium (RPE) layer, using the early treatment diabetic retinopathy study (ETDRS) grid at 1, 3 mm. Enhanced depth imaging (EDI)-OCT was used to measure subfoveal choroidal thickness (SFCT) and at 1000 µm nasal and temporal to the foveola.

Statistical Analysis

IBM SPSS 26.0 package program was used in the statistical analysis of the study. For comparisons of parameters, the independent samples t-test was used when the data were normally distributed, and the Mann–Whitney U test was used when the data were not normally distributed. Before all comparisons, the Shapiro Wilk test was used to determine whether the data were normally distributed according to groups. ANCOVA analysis was performed to account for the imbalance in gender distribution. All statistical analyzes were evaluated at the 95% confidence interval and significance at the $p<0.05$ level.

RESULTS

The trained group included 36 participants, while the untrained group included 35 participants. Thirteen participants (36.1%) of the trained group were women, while 23 (63.9%) were men. In the untrained group, 31 participants (88.6%) were women, and 4 (11.4%) were men. The average age of those who exercised was 28.42±9.10 years, while it was 29.06±6.60 years for those who did not exercise. There was no significant difference in the mean IOP (15.80±2.17 mmHg in the trained group; 15.29±1.92 mmHg in the untrained group) ($p>0.05$). A significant difference was observed in the spherical equivalent between the two groups (-0.12±0.55 in the trained group; -0.58±1.16 in the untrained group) ($p<0.05$).

There was no significant difference in BMI, HR, SBP, and DBP between the two groups (**Table 1**). All retinal layers, except for the RPE, showed no statistically significant differences ($p>0.05$) in thickness at the ETDRS grid 1-and 3-mm measurements between the trained and untrained groups (**Table 2, 3**). In the RPE layer, a significant difference was observed at the superior 3-mm area ($p=0.032$) (**Table 3**). Choroidal thickness (CT) was found to be increased in the trained group compared to the untrained group in all measurements, but only the nasal and subfoveal values were statistically significant ($p=0.013$ and $p=0.004$) (**Table 4**). After statistically adjusting for the difference in gender ratio, the significance of CT remained unchanged.

Red blood cell count (RBC) ($p=0.008$), hemoglobin (HGB) ($p=0.0001$), hematocrit (HCT) ($p=0.0001$), mean corpuscular volume (MCV) ($p=0.021$), mean corpuscular hemoglobin (MCH) ($p=0.043$) and eosinophil counts ($p=0.010$) were

Table 1. Blood pressure, heart rate, body mass index and metabolic equivalent of task comparison between groups

	Trained group n=36			Untrained group n=35			
	Median	Min-max	Mean±SD	Median	Min-max	Mean±SD	p
Systolic BP (mmHg)	114	80-130	111.44±9.65	110	90-134	109.37±11.8	0.13[a]
Diastolic BP (mmHg)	70	52-90	70.11±8.25	70	59-87	70.11±6.34	0.926[a]
Heart rate (bpm)	74.5	50-103	75.67±12.03	74	61-110	75.89±10.03	0.934[b]
BMI (kg/m^2)	24.72	19.35-33.22	24.3±2.87	22.49	14.84-45.37	23.88±4.9	0.25[a]
MET	2772	1542-9030	3727.22±2120.6	990	330-1498	991.14±335.1	0.0001[a]**

[a] Mann Whitney U test, [b] Independent samples t-test, p<0.05 *, p<0.01**, SD: Standard deviation, Min: Minimum, Max: Maximum, Systolic BP: Systolic blood pressure, Diastolic BP: Diastolic blood pressure, BMI: Body-mass index, MET: Metabolic equivalent of task

Table 2. Thickness of all retinal layers in central 1-mm of ETDRS grid

	Trained group n=36			Untrained group n=35			
	Median	Min-max	Mean±SD	Median	Min-max	Mean±SD	p
RT	262.5	219-346	264.67±23.8	264	225-317	265.26±18.1	0.645[a]
RNFL	12	6-20	12.31±2.6	12	8-23	12.34±2.6	0.74[a]
GCL	15.5	8-52	16.94±8.5	14	5-36	15.66±6	0.624[a]
IPL	21	12-42	21.25±5.9	20	14-30	20.14±3.6	0.607[a]
INL	17	10-38	18.08±5	17	9-33	17.8±5.1	0.782[a]
OPL	23	12-32	23.03±4.7	25	16-54	25.91±7	0.057[a]
ONL	87.5	66-106	86.44±9.3	89	43-112	86.83±13.1	0.887[b]
RPE	17	13-22	17.31±2	17	13-21	17.2±2.3	0.837[b]

[a] Mann Whitney U test, [b] Independent samples t-test, p<0.05 *, p<0.01**, SD: Standard deviation, Min: Minimum, Max: Maximum, ETDRS: Early Treatment Diabetic Retinopathy Study, RT: Retinal thickness, RNFL: Retinal nerve fiber layer thickness, GCL: Ganglion cell layer, IPL: Inner plexiform layer, OPL: Outer plexiform layer, ONL: Outer nuclear layer, RPE: Retinal pigment epithelial layer. Thicknesses are expressed in micrometers. p<0.05 shows statistical significance.

statistically significantly higher in the trained group (**Figure 1**). Statistical subgroup analysis was performed considering the difference in gender ratio. In women, only white blood cell values showed a significant difference, whereas in men, HGB, HCT, and HDL levels were significantly higher in the trained group (p>0.05), while triglyceride levels were significantly higher in the untrained group (p<0.05) (**Figure 2a, b**).

Participants engaged in regular exercise for a mean period of 4.29±2.5 years. Exercise duration per week was calculated according to intensity. The mean weekly duration of moderate-intensity exercise (4 METs) averaged 389.31±383.70 minutes, while high-intensity exercise (8 METs) was 271.81±218.10 minutes. The mean length of a single exercise session was 73.71±55.10 minutes for moderate-intensity exercise and 76.29±53.20 minutes for high-intensity exercise.

The weekly duration of high-intensity exercise showed a positive correlation with RBC, HGB, HCT, and MCH. Furthermore, the length of individual high-intensity exercise sessions was positively correlated with HGB, HCT, MCV, and MCH. The total MET value demonstrated a positive correlation with SFCT, whereas no correlation was observed between CT and RPE thickness.

DISCUSSION

This study is the first to evaluate the effect of regular exercise during 4 years on the thickness of the choroid and retina layers, BP, hematological and lipid profiles.

Increases in HR and SBP correlates with OPP during exercise.[10] Choroidal, retinal, and optic nerve head blood flow tends to remain stable with increases in OPP of up to 40–60%. In an isometric exercise study, a 60% rise in OPP did not elevate CBF, suggesting regulatory mechanisms of CBF.[11] Lovasik et al.[12] examined the relationship between OPP and CBF during a 20-minute cycling protocol and at rest. OPP increased by 43% at exercise onset, subsequently decreasing to approximately 10% above resting levels at the end of the test.[12] In contrast, CBF demonstrated a linear increase and remained within 10% of baseline values.[12] These findings suggest that CBF is primarily regulated by arteriolar sympathetic vasoconstriction during exercise.[12]

An OCT-angiography study reported a decrease in vessel density at the superficial capillary plexus, which correlated with increased SBP during exercise.[3] Similarly, Szalai et al.[1] observed retinal thinning one minute after intense exercise, followed by retinal thickening at 5 and 15 minutes post-exercise. They attributed this result to autoregulatory vasoconstriction during exercise and vasodilation post-exercise.[1] In the present study, significant RPE thinning was only observed in the 3 mm superior subfield of the ETDRS grid. The RPE thinning observed in this study may be secondary to the enlargement of choroidal vessels and compression by a thickened choriocapillaris, resembling pachychoroid morphology.[13] However, this finding could also represent an incidental observation. To date, only one study has evaluated the long-term effects of exercise on the retina, specifically measuring retinal vessel diameter in marathon runners, and found no significant changes.[14]

Different results have been reported in studies investigating the effect of exercise on the choroid. Alwassia et al.[15] did

Table 3. Thickness of all retinal layers in inner 3-mm ring of ETDRS grid

	Trained group n=36			Untrained group n=35			
	Median	Min-max	Mean±SD	Median	Min-max	Mean±SD	p
RT							
3-S	339	307-366	339.33±15.1	343	313-360	340.91±12.9	0.625[a]
3-I	339	302-373	338.42±16.5	339	306-367	340.4±15.5	0.604[b]
3-N	337.5	251-364	337.67±19.7	340	293-368	339.37±15.7	0.809[a]
3-T	327	300-358	326.25±14.6	328	12-350	316.66±55.2	0.927[a]
RNFL							
3-S	23	17-29	22.61±2.7	23	18-29	23.17±2.7	0.384[b]
3-I	25.5	14-30	25.36±3.5	27	18-38	26.63±3.7	0.301[a]
3-N	20	16-27	20.64±2.3	20	15-27	20.74±2.5	0.802[a]
3-T	16	15-19	16.44±1	17	14-20	16.86±1.3	0.121[a]
GCL							
3-S	53	41-63	52.75±5	52	44-65	53.11±4.8	0.755[b]
3-I	53.5	28-66	53±6.4	53	45-61	53.74±3.8	0.632[a]
3-N	53	43-68	53.11±5.3	52	37-61	51.54±4.9	0.203[b]
3-T	49	38-59	48.47±5.3	48	27-56	47.2±5.7	0.381[a]
IPL							
3-S	41	33-48	41.14±3.3	42	36-48	41.66±3.1	0.503[b]
3-I	41	31-50	41.25±3.7	42	34-48	42.09±3.2	0.315[b]
3-N	42	36-50	41.89±3.5	43	32-48	41.86±3.6	0.970[b]
3-T	41	32-50	41.08±3.8	41	27-46	40.69±3.8	0.954[a]
INL							
3-S	41	33-55	40.92±5	41	35-46	40.17±3.4	0.463[b]
3-I	40	33-47	40.64±3.8	40	34-46	40.57±3.2	0.935[b]
3-N	41.5	34-46	40.89±3.4	41	31-47	40.03±3.5	0.274[a]
3-T	37	30-44	37.42±3.6	38	26-45	37.49±3.9	0.939[b]
OPL							
3-S	30.5	25-57	33.64±9.1	29	23-55	31.89±7.5	0.413[a]
3-I	28	24-53	30.83±6.3	32	24-51	33.03±6.5	0.050[a]
3-N	31	26-40	31.42±4	31	24-63	32.57±8.3	0.817[a]
3-T	29	25-39	30.08±3.3	29	24-44	29.6±4.6	0.364[a]
ONL							
3-S	68.5	45-84	67.67±10	72	46-86	69.89±11.1	0.191[a]
3-I	70.5	36-85	68.06±10.5	64	46-79	64.09±9.6	0.072[a]
3-N	72	56-85	70.72±8	70	38-88	69.94±12	0.778[a]
3-T	71,5	52-85	71.33±6.9	73	50-86	71.86±8.3	0.496[a]
RPE							
3-S	15	13-19	14.89±1.2	15	13-18	15.37±1.3	0.032[a*]
3-I	14.5	12-19	14.53±1.4	14	12-17	14.2±1.3	0.437[a]
3-N	15	12-18	14.97±1.2	15	12-18	15.34±1.5	0.182[a]
3-T	15	13-17	14.58±1.1	14	11-17	14.26±1.4	0.381[a]

[a] Mann Whitney U test, [b] Independent samples t-test, p<0.05 *, p<0.01**, SD: Standard deviation, Min: Minimum, Max: Maximum, ETDRS: Early Treatment Diabetic Retinopathy Study, 3S: 3 mm superior subfield, 3I: 3 mm inferior subfield, 3N: 3 mm nasal subfield, 3T: 3 mm temporal subfield, RT: Retinal thickness, RNFL: Retinal nerve fiber layer thickness, GCL: Ganglion cell layer, IPL: Inner plexiform layer, OPL: Outer plexiform layer, ONL: Outer nuclear layer, RPE: Retinal pigment epithelial layer. Thicknesses are expressed in micrometers. p<0.05 shows statistical significance.

not find any change in CT at the 3rd minute of stress testing in patients with a mean age of 60 years. Kinoshita et al.[16] analyzed the choroid more thoroughly after mild dynamic exercise. They observed no change in central CT, mean luminal and stromal areas, or the mean luminal/choroidal area ratio within 10 minutes after exercise.[16] In contrast, the study by Sayın et al.[17] found thickening of the choroid 5 minutes after medium-intensity training, which returned to normal values 15 minutes post-exercise. In the present study, temporal, subfoveal, and nasal CT was found to be increased in the trained group compared to the untrained group, but only the nasal and subfoveal measurements were statistically

Table 4. Choroidal thickness in nasal, subfoveal and temporal measurements

	Trained group n=36			Untrained group n=35			
	Median	Min-max	Mean±SD	Median	Min-max	Mean±SD	p
Nasal CT	379	226-584	380.22±90	342	197-458	333.31±63.5	0.013[a*]
Subfoveal CT	396.5	267-619	405.08±84.6	355	168-483	348.8±74.6	0.004[a**]
Temporal CT	384.5	249-554	383.36±88.3	344	243-464	349.51±55.1	0.057[a]

* Independent samples t-test, p<0.05 *, p<0.01**, SD: Standard deviation, Min: Minimum, Max: Maximum, CT: Choroidal thickness. Thicknesses are expressed in micrometers.

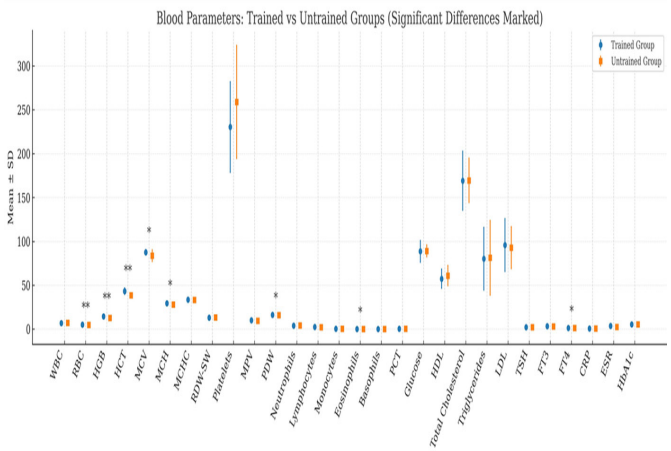


Figure 1. Comparison of hematological and biochemical parameters between trained and untrained groups. Overall analysis including both sexes.
Data are presented as mean±SD. Blue circles represent the trained group, and red squares represent the untrained group. Asterisks denote statistically significant differences (*p<0.05, **p<0.01; Independent samples t-test, Mann–Whitney U test). WBC: White blood cell, RBC: Red blood cell, HGB: Hemoglobin, HCT: Hematocrit, MCV: Mean corpuscular volume, MCH: Mean corpuscular hemoglobin, MCHC: Mean corpuscular hemoglobin concentration, RDW-SW: Red blood cell distribution width-standard deviation, MPV: Mean platelet volume, PDW: Platelet distribution width, PCT: Procalcitonin, HDL: High-density lipoprotein, LDL: Low-density lipoprotein, TSH: Thyroid-stimulating hormone, CRP: C-reactive protein, ESR: Erythrocyte sedimentation rate

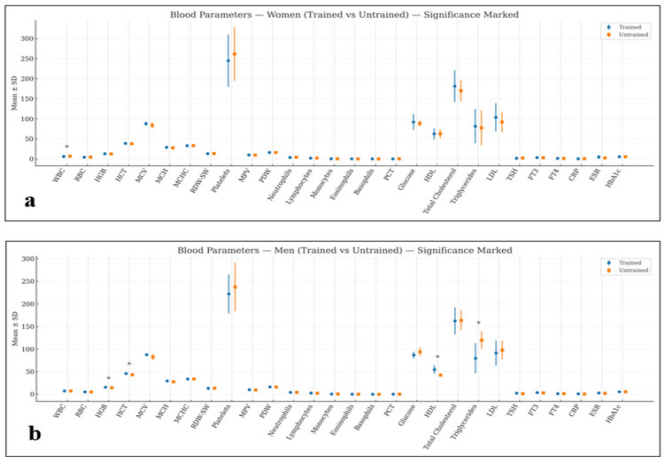


Figure 2. Hematological and biochemical parameters in trained versus untrained groups, stratified by sex: **a)** women **b)** men
Data are presented as mean±SD. Blue circles represent the trained group, and red squares represent the untrained group. Asterisks denote statistically significant differences (*p<0.05, **p<0.01; Mann–Whitney U test). WBC: White blood cell, RBC: Red blood cell, HGB: Hemoglobin, HCT: Hematocrit, MCV: Mean corpuscular volume, MCH: Mean corpuscular hemoglobin, MCHC: Mean corpuscular hemoglobin concentration, RDW-SW: Red blood cell distribution width-standard deviation, MPV: Mean platelet volume, PDW: Platelet distribution width, PCT: Procalcitonin, HDL: High-density lipoprotein, LDL: Low-density lipoprotein, TSH: Thyroid-stimulating hormone, CRP: C-reactive protein, ESR: Erythrocyte sedimentation rate

significant (p<0.05 and p<0.01). There is only one long-term study (the CHAMPS study) in the literature that aimed to find a correlation between physical activity and CT in children.[18] They reported no association between CT and physical activity.[18]

Differences in exercise study outcomes may be attributed to variations in exercise duration, intensity, and the occurrence of hypocapnia. CBF increases by more than 10% when OPP rises by over 60%, depending on exercise intensity and duration.[11,12] Intense exercise could lead to hyperventilation and hypocapnia.[19] A study showed that retinal and CBF increased during submaximal exercise at 6 minutes but decreased with exhaustion due to hypocapnia.[20] In the present study, the observed increase in CT suggests that CBF rose to a level sufficient to induce thickening of the choroid under the given exercise duration and intensity.

The spherical equivalent was significantly more myopic in the untrained group (-0.12±0.55 in the trained group; -0.58±1.16 in the untrained group) (p<0.05). An increase of one diopter in myopia has been shown to decrease SFCT by 13 µm.[21] SFCT was 405.08±84.6 µm in the trained group and 348.8±74.6 µm in the untrained group. Even after recalculating, considering a -0.50-diopter difference, SFCT in the trained group would still be statistically significantly thicker.

In the present study, RBC, HGB, HCT, MCV, MCH, and eosinophil values were observed to be significantly higher in the trained group. However, no correlation was identified between these hematological parameters and CT. Studies on hematological parameters show varying results depending on exercise intensity and whether the exercise is anaerobic or aerobic. A study on a group that regularly engaged in aerobic exercise over the past 12 weeks showed increased HGB, HCT, and platelet levels.[22] Another study compared the effects of aerobic and strengthening exercises after 16 weeks of training and reported significant decreases in RBC, HGB, HCT, and MCV values in the strengthening exercise group.[23] Consistent with these findings, Alam et al.[24] also observed similar hematologic changes in athletes following intense exercise. Bizjak et al.[25] evaluated the effects of a 6-week moderate training protocol on RBCs and found an increase in both young and old RBCs. They attributed the RBC results in their study to the increase in catecholamines, cortisol, growth hormones, and insulin-like factors with exercise, which stimulate erythropoiesis.[25]

Limitations

One of the limitations of the study is the difference in the female/male ratio between the groups. Some studies have found that the choroid is thicker in males than in females, while others have found no difference.[26,27] Another limitation is the absence of an assessment of patients' dietary habits and

micronutrient intake, factors that may significantly influence hematological and biochemical parameters.

The elevated eosinophil count observed in this study may reflect non-allergic activation.[28,29] Several studies showed that eosinophils could be activated by exercise in healthy individuals, with both long-term endurance and short-term maximal exercise elevating eosinophil cationic protein levels.[28,29]

In this study, no significant differences were observed in BMI, HR, SBP, DBP, lipid profile, HbA1c, or thyroid function between the trained and untrained groups. A study investigating six months of moderate- and high-intensity exercise reported significant changes in total cholesterol, LDL, and HDL levels in the high-intensity group.[30] Consistent with these findings, our subgroup analysis in men had higher levels of HDL. These findings suggest that both the duration and type of training are key determinants of exercise-related effects on lipid profiles and BP, consistent with the other parameters evaluated in the present study.[30,31] The training regimen of the study group does not appear to have been sufficient to alter lipid profiles and BP in the present study.

CONCLUSION

As a result, to the best of our knowledge, this is the first study to evaluate choroidal and retinal parameters, along with hematologic and lipid profiles, in a predominantly moderately active population engaged in long-term exercise under real-life conditions. CT, RBC, HGB, HCT, MCV, MCH, and eosinophil values were statistically significantly increased in the trained group under real-life conditions. The characteristics of exercise—such as duration, intensity, endurance, and whether it is aerobic or anaerobic—may be the main factors determining the systemic and organ effects of exercise.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was approved by the University of Health Sciences Hamidiye Clinical Researches Ethics Committee (Date: 23.11.2023, Decision No: 2023.23.11-63).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Szalai I, Csorba A, Pálya F, et al. The assessment of acute chorioretinal changes due to intensive physical exercise in young adults. *PLoS One*. 2022;17(5):e0268770. doi:10.1371/journal.pone.0268770
2. Risner D, Ehrlich R, Kheradiya NS, Siesky B, McCranor L, Harris A. Effects of exercise on intraocular pressure and ocular blood flow: a review. *J Glaucoma*. 2009;18(6):429-436. doi:10.1097/IJG.0b013e31818fa5f3
3. Vo Kim S, Semoun O, Pedinielli A, Jung C, Miere A, Souied EH. Optical coherence tomography angiography quantitative assessment of exercise-induced variations in retinal vascular plexa of healthy subjects. *Invest Ophthalmol Vis Sci*. 2019;60(5):1412-1419. doi:10.1167/iovs.18-24389
4. Luo X, Shen YM, Jiang MN, Lou XF, Shen Y. Ocular blood flow autoregulation mechanisms and methods. *J Ophthalmol*. 2015;2015:864871. doi:10.1155/2015/864871
5. Yeung SC, You Y, Howe KL, Yan P. Choroidal thickness in patients with cardiovascular disease: a review. *Surv Ophthalmol*. 2020;65(4):473-486. doi:10.1016/j.survophthal.2019.12.007
6. Yumusak E, Ciftci A, Yalcin S, Sayan CD, Dikel NH, Ornek K. Changes in the choroidal thickness in reproductive-aged women with iron-deficiency anemia. *BMC Ophthalmol*. 2015;15(1):186. doi:10.1186/s12886-015-0163-3
7. Fleenor BS, Marshall KD, Durrant JR, Lesniewski LA, Seals DR. Arterial stiffening with ageing is associated with transforming growth factor-β1-related changes in adventitial collagen: reversal by aerobic exercise. *J Physiol*. 2010;588(Pt 20):3971-3982. doi:10.1113/jphysiol.2010.194753
8. Nystoriak MA, Bhatnagar A. Cardiovascular effects and benefits of exercise. *Front Cardiovasc Med*. 2018;5(135):11. doi:10.3389/fcvm.2018.00135
9. Pedersen BK, Saltin B. Exercise as medicine-evidence for prescribing exercise as therapy in 26 different chronic diseases. *Scand J Med Sci Sports*. 2015;25(Suppl 3):1-72. doi:10.1111/sms.12581
10. Gale J, Wells AP, Wilson G. Effects of exercise on ocular physiology and disease. *Surv Ophthalmol*. 2009;54(3):349-355. doi:10.1016/j.survophthal.2009.02.005
11. Popa-Cherecheanu A, Schmidl D, Werkmeister RM, Chua J, Garhöfer G, Schmetterer L. Regulation of choroidal blood flow during isometric exercise at different levels of intraocular pressure. *Invest Ophthalmol Vis Sci*. 2019;60(1):176-182. doi:10.1167/iovs.18-24992
12. Lovasik JV, Kergoat H, Riva CE, Petrig BL, Geiser M. Choroidal blood flow during exercise-induced changes in the ocular perfusion pressure. *Invest Ophthalmol Vis Sci*. 2003;44(5):2126-2132. doi:10.1167/iovs.02-0825
13. Baek J, Kook L, Lee WK. Choriocapillaris flow impairments in association with pachyvessel in early stages of pachychoroid. *Sci Rep*. 2019;9(1):5565. doi:10.1038/s41598-019-42052-w
14. Pressler A, Hanssen H, Dimitrova M, Krumm M, Halle M, Scherr J. Acute and chronic effects of marathon running on the retinal microcirculation. *Atherosclerosis*. 2011;219(2):864-868. doi:10.1016/j.atherosclerosis.2011.08.052
15. Alwassia AA, Adhi M, Zhang JY, et al. Exercise-induced acute changes in systolic blood pressure do not alter choroidal thickness as measured by a portable spectral-domain optical coherence tomography device. *Retina*. 2013;33(1):160-165. doi:10.1097/IAE.0b013e3182618c22
16. Kinoshita T, Mori J, Okuda N, et al. Effects of exercise on the structure and circulation of choroid in normal eyes. *PLoS One*. 2016;11(12):e0168336. doi:10.1371/journal.pone.0168336
17. Sayin N, Kara N, Pekel G, Altinkaynak H. Choroidal thickness changes after dynamic exercise as measured by spectral-domain optical coherence tomography. *Indian J Ophthalmol*. 2015;63(5):445-450. doi:10.4103/0301-4738.159884
18. Lundberg K, Vestergaard AH, Jacobsen N, et al. Choroidal thickness and myopia in relation to physical activity-the CHAMPS eye study. *Acta Ophthalmol*. 2018;96(4):371-378. doi:10.1111/aos.13640
19. Tsakiris TS, Konstantopoulos AI, Bourdas DI. The role of CO_2 on respiration and metabolism during hypercapnic and normocapnic recovery from exercise. *Res Q Exerc Sport*. 2021;92(3):537-548. doi:10.1080/02701367.2020.1756196

20. Ikemura T, Hayashi N. Ocular circulatory responses to exhaustive exercise in humans. *Eur J Appl Physiol*. 2012;112(9):3313-3318. doi:10.1007/s00421-012-2313-0
21. Zhu H, Liu C, Gao M, Zhang S, Zhang L, Zhao Q. Choroidal thickness in relation to diopter and axial length among myopic children. *Front Med (Lausanne)*. 2023;10:1241352. doi:10.3389/fmed.2023.1241352
22. Wardyn GG, Rennard SI, Brusnahan SK, et al. Effects of exercise on hematological parameters, circulating side population cells, and cytokines. *Exp Hematol*. 2008;36(2):216-223. doi:10.1016/j.exphem.2007.10.003
23. Cicek G. The effects of different exercise types on hematological parameters in sedentary women. *J Educ Training Studies*. 2018;6(8):96. doi:10.11114/jets.v6i8.3374.
24. Alam T, Rahman SM, Alam T, et al. Effect of physical exercise on some hematological parameters in female athletes in Bangladesh. *JNMA J Nepal Med Assoc*. 2014;52(195):892-896.
25. Bizjak DA, Tomschi F, Bales G, et al. Does endurance training improve red blood cell aging and hemorheology in moderate-trained healthy individuals? *J Sport Health Sci*. 2020;9(6):595-603. doi:10.1016/j.jshs.2019.02.002
26. Li XQ, Larsen M, Munch IC. Subfoveal choroidal thickness in relation to sex and axial length in 93 Danish university students. *Invest Ophthalmol Vis Sci*. 2011;52(11):8438-8441. doi:10.1167/iovs.11-8108
27. Okonkwo ON, Hassan AO, Oyekunle I, Akanbi T, Chinedu UB. Choroidal thickness and effect of sex and age in Africans. *Ann Afr Med*. 2023;22(4):489-496. doi:10.4103/aam.aam_184_22
28. Domej W, Schwaberger G, Tilz GP, et al. Prolonged endurance challenge at moderate altitude: effect on serum eosinophil cationic protein, eosinophil dynamics, and lung function. *Chest*. 2002;121(4):1111-1116. doi:10.1378/chest.121.4.1111
29. Dufaux B, Heine O, Prinz U, Soll C, Rost R. Effect of a short maximal physical exercise on the eosinophil cationic protein. *Int J Sports Med*. 1993;14(8):468-470. doi:10.1055/s-2007-1021213
30. O'Donovan G, Owen A, Bird SR, et al. Changes in cardiorespiratory fitness and coronary heart disease risk factors following 24 wk of moderate- or high-intensity exercise of equal energy cost. *J Appl Physiol (1985)*. 2005;98(5):1619-1625. doi:10.1152/japplphysiol.01310.2004
31. Cardoso CG Jr, Gomides RS, Queiroz AC, et al. Acute and chronic effects of aerobic and resistance exercise on ambulatory blood pressure. *Clinics (Sao Paulo)*. 2010;65(3):317-325. doi:10.1590/S1807-59322010000300013

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1759128

Anatolian Curr Med J. 2025;7(5):642-646

Perinatal outcomes of isolated non-visualization of the fetal gallbladder

Ayşe Çiğdem Bayrak, Recep Taha Ağaoğlu, Berna Seyhan, Özgür Volkan Akbulut, Aysu Yeşim Tezcan, Kadriye Yakut Yücel

Department of Obstetrics and Gynecology, Ankara Etlik City Hospital, Ankara, Turkiye

Cite this article as: Bayrak AÇ, Ağaoğlu RT, Seyhan B, Akbulut ÖV, Tezcan AY, Yakut Yücel K. Perinatal outcomes of isolated non-visualization of the fetal gallbladder. *Anatolian Curr Med J.* 2025;7(5):642-646.

Received: 06.08.2025 • **Accepted**: 01.09.2025 • **Published**: 15.09.2025

ABSTRACT

Aims: This study aimed to evaluate perinatal outcomes in pregnancies with isolated non-visualization of the fetal gallbladder (NVFGB) identified during second-trimester anatomical screening.

Methods: This retrospective cohort study included 22 pregnancies diagnosed with isolated NVFGB between November 2022 and January 2025 at a tertiary maternal–fetal medicine unit. Cases with additional structural anomalies, multiple gestations, or elective terminations were excluded. Maternal demographics, antenatal ultrasound findings, and neonatal outcomes were reviewed. Postnatal imaging and clinical follow-up were evaluated for gallbladder visualization and underlying pathology.

Results: In half of the included cases (11/22), the gallbladder was visualized either on follow-up scans or after birth. Among the remaining 11 cases, two had midline-located gallbladders on postnatal imaging, one of which required surgical correction for intestinal malrotation. Two fetuses were prenatally diagnosed with cystic fibrosis, including one complicated by meconium peritonitis requiring surgery. Two additional cases were diagnosed postnatally with biliary atresia and underwent hepatoportoenterostomy. Chromosomal microarray analysis (CMA) was performed in ten cases; no anomalies were identified aside from cystic fibrosis. Overall, five cases (23%) were associated with significant postnatal diagnoses requiring medical or surgical intervention.

Conclusion: Although isolated NVFGB is often a benign and transient finding, it may occasionally indicate serious underlying conditions such as biliary atresia or cystic fibrosis. Detailed follow-up, repeat imaging in late gestation, and thorough postnatal evaluation are essential for appropriate diagnosis and management.

Keywords: Biliary tract abnormalities, gallbladder abnormalities, prenatal ultrasonography

INTRODUCTION

The gallbladder originates from the hepatic diverticulum during early embryogenesis, and disruptions in this process may lead to anomalies such as agenesis, biliary atresia, or ectopic localization.[1,2] Among these, gallbladder agenesis is relatively uncommon, with a reported prevalence of approximately 0.1%.[3] Nevertheless, its occurrence represents a significant prenatal finding due to its potential association with a variety of structural and genetic conditions—including other gastrointestinal anomalies, cardiac malformations, and chromosomal abnormalities. Consequently, non-visualization of the gallbladder on prenatal imaging necessitates a thorough anatomical assessment. In selected cases, evaluation of digestive enzyme levels in the amniotic fluid, in addition to invasive genetic testing, may also provide valuable information.[4,5]

Although non-visualization of the fetal gallbladder is frequently reported as a prenatal ultrasound finding during early second-trimester sonography, subsequent imaging—either later in gestation or postnatally—often confirms the presence of a normally developed gallbladder. In approximately one-quarter of these cases, the gallbladder becomes visible on follow-up ultrasounds. Moreover, in up to 40% of fetuses with persistent non-visualization throughout pregnancy, postnatal imaging reveals a structurally normal gallbladder.[6] This discrepancy highlights the potential for false-positive diagnoses and underscores the dynamic nature of fetal gallbladder visualization. While most isolated cases are associated with favorable outcomes, the possibility of underlying structural or genetic abnormalities necessitates careful differential diagnosis.[7]

In this context, the present study aimed to evaluate the ultrasonographic characteristics and perinatal outcomes of cases with isolated non-visualization of the fetal gallbladder (NVFGB) detected during second-trimester anatomical screening.

Corresponding Author: Ayşe Çiğdem Bayrak, drcigdembayrak@gmail.com

METHODS

The study population was identified through a review of electronic medical records of pregnant women with NVFGB during second-trimester anatomical screening. All fetal anatomical screening records between November 1, 2022, and January 1, 2025, were reviewed. The study protocol was approved by the Scientific Researches Evaluation and Ethics Committee No. 1 at Ankara Etlik City Hospital (Date: 14.05.2025, Decision No: AESH-BADEK1-2025-204) and conducted in accordance with the Declaration of Helsinki.

In this study, NVFGB was defined as the inability to identify the gallbladder on second-trimester anatomical ultrasound and on a confirmatory subsequent scan, despite adequate visualization of other abdominal organs. Only patients who received both antenatal follow-up and delivery at our institution were included. For isolated NVFGB, follow-up sonographic data were reviewed to assess gallbladder visibility and the gestational age at which visualization occurred. When the gallbladder remained non-visualized antenatally, postnatal follow-up and neonatal imaging, when available, were used to distinguish transient non-visualization from true agenesis. In cases requiring surgical intervention, the type and timing of procedures were recorded to provide further insight into the clinical significance and outcomes of gallbladder agenesis.

Pregnancies complicated by additional structural anomalies, multiple gestations, or terminated due to fetal or maternal indications were excluded. Maternal demographic characteristics, results of prenatal screening tests, follow-up sonographic findings, and mode of delivery were recorded. Neonatal outcomes, including gestational age at delivery, birth weight percentile, 1- and 5-minute Apgar scores, admission to the neonatal intensive care unit (NICU), and any postnatal imaging or endoscopic procedures related to the gallbladder, were also evaluated.

Statistical Analysis

Data were analyzed using IBM SPSS Statistics for Windows, Version 23.0 (IBM Corp., Armonk, NY, USA). Continuous variables were presented as mean±standard deviation or median (minimum–maximum), and categorical variables as counts and percentages. Only descriptive statistics were presented; no group comparisons were made.

RESULTS

This study included 22 cases of isolated NVFGB identified during second-trimester anatomical screening, with a mean gestational age at diagnosis of 18+6 weeks. Maternal and neonatal baseline characteristics are summarized in **Table 1**. In terms of neonatal outcomes, six neonates (27%) required admission to the NICU, including one case in which the gallbladder was later visualized and five cases with clinically significant postnatal diagnoses in addition to gallbladder agenesis.

Overall, the presence of the gallbladder was confirmed in 11 of the 22 cases (50%), either antenatally or postnatally.

Table 1. Maternal characteristics and neonatal outcomes (n=22)

Maternal age*	31.32 (24-39)
Gravidity*	1.73 (1-4)
Parity*	0.5 (0-2)
Abortus*	0.5 (0-2)
Parental consanguinity†	
Yes	2 (9.09%)
No	20 (90.9%)
Gestational age at ultrasound*	18+6 (15-22)
Birth week*	38+3 (36-40+3)
Birth weight (g)*	3128.6 (2410-3680)
Apgar score 1st min*	7.45 (4-9)
Apgar score 5th min*	8.68 (5-10)
Blood gas pH value*	7.30 (7.05-7.46)
NICU admission†	6 (27.2%)

*Values are presented as median (minimum–maximum); †Values are presented as number (percentage). NICU: Neonatal intensive care unit

Specifically, in eight cases, the gallbladder became visible on follow-up scans performed in the late second or early third trimester. In another four cases, a thin, hyperechoic tubular structure was noted in the expected location of the gallbladder, although the typical morphology was absent; postnatal sonography confirmed the presence of the gallbladder in three of these. In the remaining case, postnatal imaging failed to demonstrate a definitive gallbladder, and the appearance was interpreted as consistent with agenesis. The outcomes of these cases are detailed in **Table 2**.

In the remaining 11 cases, the gallbladder could not be identified in its expected location on postnatal imaging. Among these, two cases demonstrated atypical findings, with the gallbladder located in the midline of the abdomen. One of these was subsequently diagnosed with intestinal malrotation and underwent surgical intervention on postnatal day 5 due to suspected intestinal obstruction. During laparotomy, extensive Ladd's bands were excised, ileal loops were mobilized, and the gallbladder was visualized. The other case remained asymptomatic, with no abnormalities detected during seven months of postnatal follow-up.

Two fetuses had a prenatal diagnosis of cystic fibrosis, one of whom was born to consanguineous parents. One developed meconium peritonitis and required ileal resection with primary anastomosis on postnatal day 1. Biliary atresia was identified in two further cases—one presenting antenatally with a hyperechoic tubular structure and the other with complete non-visualization—both of which underwent hepatoportoenterostomy (Kasai procedure) on postnatal days 6 and 11, respectively.

Chromosomal microarray analysis (CMA) was performed in six cases; aside from the two with cystic fibrosis, no anomalies were detected. In total, five cases (23%) were associated with clinically significant postnatal diagnoses requiring medical or surgical intervention. These findings are summarized in **Table 3**.

Table 2. Outcomes of fetuses with NVFGB and subsequent visualization on follow-up or postnatal evaluation (n=11)

Case no	GA at 1st USG (wks)	Antenatal GB visibility	GA at GB visualization	Genetic testing	Postnatal GB
1	17	NVFGB	24	Not performed	Normal
2	19	NVFGB	28	Not performed	Normal
3	19	NVFGB	27	Not performed	Normal
4	21	NVFGB	24	Not performed	Normal
5	20	NVFGB	29	Normal CMA	Normal
6	20	NVFGB	31	Not performed	Normal
7	19	NVFGB	33	Normal CMA	Normal
8	22	NVFGB	26	Not performed	Normal
9	20	Hyperechoic tubular structure	Not visualized	Normal CMA	Normal
10	18	Hyperechoic tubular structure	Not visualized	Normal CMA	Normal
11	19	Hyperechoic tubular structure	Not visualized	Not performed	Normal

GA: Gestational age, USG: Ultrasonography, GB: Gallbladder, NVFGB: Non-visualization of the fetal gallbladder, CMA: Chromosomal microarray analysis

Table 3. Outcomes of fetuses with persistent NVFGB diagnosed with gallbladder agenesis (n=11)

Case no	GA at 1st USG (wks)	Antenatal GB visibility	Genetic testing	Postnatal diagnosis	Postnatal course
1	19	NVFGB	Not performed	Contracted GB	Asymptomatic
2	19	NVFGB	Normal CMA	Midline GB	Asymptomatic
3	20	NVFGB	Normal CMA	Midline GB+intestinal malrotation	Laparatomy
4	17	NVFGB	Cystic fibrosis	GB Agenesis+meconium peritonitis	Ileal resection and primary anastomosis
5	16	NVFGB	Cystic fibrosis	GB agenesis	Asymptomatic
6	17	NVFGB	Not performed	GB agenesis	Asymptomatic
7	19	NVFGB	Normal CMA	GB agenesis	Asymptomatic
8	15	NVFGB	Not performed	GB agenesis	Asymptomatic
9	18	NVFGB	Normal CMA	GB agenesis	Asymptomatic
10	21	NVFGB	Normal CMA	GB agenesis+biliary atresia	Hepatoportoenterostomy
11	19	Hyperechoic tubular structure	Normal CMA	GB Agenesis+biliary atresia	Hepatoportoenterostomy

GA: Gestational age, USG: Ultrasonography, GB: Gallbladder, NVFGB: Non-visualization of the fetal gallbladder, CMA: Chromosomal microarray analysis

DISCUSSION

NVFGB during second-trimester sonography is generally regarded as a benign finding in isolated cases, as the gallbladder is often visualized later in gestation or confirmed postnatally.[8,9] In our study, approximately half of the fetuses with either prenatal non-visualization or a small, hyperechoic tubular appearance of the gallbladder were confirmed to have a gallbladder on postnatal imaging. In eight cases, the gallbladder became visible in subsequent prenatal scans, and no postnatal abnormalities were identified. These findings suggest the possibility that isolated non-visualization may reflect physiological variability or technical limitations rather than true pathology. Therefore, serial ultrasound evaluations during pregnancy are recommended in such cases to avoid unnecessary interventions.

Nevertheless, persistent non-visualization may sometimes represent the only prenatal indicator of significant underlying pathology. One of the most critical conditions in this context is biliary atresia—a progressive fibro-obliterative disease of the extrahepatic bile ducts that can lead to liver failure in the absence of timely intervention.[10,11] Although various anatomical variants have been described, the gallbladder is frequently absent in cases of biliary atresia; when present, it is usually abnormal in size and may be associated with hepatic hilar cysts.[12] A hallmark prenatal sonographic feature of biliary atresia is the 'triangular cord sign,' first described by Choi et al.[13] in 1996, which refers to an echogenic band ≥4 mm in thickness located anterior to the portal vein bifurcation, thought to represent fibrotic remnants of the bile ducts. Although several sonographic features—such as abnormal gallbladder morphology, non-visualization of the common bile duct, hepatic subcapsular flow, or increased hepatic artery diameter—have been associated with biliary atresia, most of these findings are typically identified postnatally, and their utility in antenatal diagnosis remains limited.[14,15]

In the literature, postnatal diagnosis of biliary atresia has been reported in approximately 4% of cases with isolated non-visualization of the gallbladder on prenatal ultrasound; however, gallbladder agenesis is observed in almost all cases of biliary atresia.[16,17] Consistent with these findings, two cases in our cohort were diagnosed with biliary atresia postnatally and subsequently underwent Kasai portoenterostomy.

In addition to hepatobiliary anomalies, gastrointestinal malformations are also commonly associated with gallbladder agenesis. In our series, one fetus was diagnosed with intestinal malrotation that required surgical intervention. Furthermore,

two cases were prenatally diagnosed with cystic fibrosis. In one of these, meconium peritonitis was detected shortly before birth, necessitating ileal resection and anastomosis on the first day of life. While uncommon, NVFGB may serve as the earliest—and sometimes the only—prenatal indicator of cystic fibrosis. In the presence of associated findings such as echogenic bowel, bowel dilation, or meconium peritonitis, genetic counseling and molecular testing for cystic fibrosis should be considered.[18,19]

The fetal gallbladder exhibits a wide range of morphological variations, including folded, septated, spherical, or angular configurations. These variants are typically benign in the absence of additional structural anomalies. When a detailed second-trimester ultrasound reveals no other abnormalities, such morphological differences are generally regarded as physiological.[9] However, in any case of suspected gallbladder agenesis, a meticulous evaluation of the intra-abdominal anatomy, particularly the liver, porta hepatis, and bile ducts, is essential to exclude underlying pathology.

This study contributes to the current literature by specifically focusing on isolated cases of gallbladder non-visualization and providing a detailed analysis of both antenatal and postnatal outcomes.

Limitations

This study has several limitations. The retrospective design may have introduced selection and information bias. The relatively small sample size may have limited the statistical power to detect rare associations. In addition, the absence of genetic testing in some cases could have led to an underestimation of underlying genetic conditions. Finally, the lack of long-term follow-up data restricts the ability to evaluate delayed postnatal outcomes.

CONCLUSION

As a result, although NVFGB during second-trimester anatomical screening is often a benign and transient finding, it may occasionally represent the earliest manifestation of serious hepatobiliary or systemic conditions. Therefore, even in isolated cases, comprehensive prenatal and postnatal evaluation, repeat imaging in later gestation, and a multidisciplinary approach are warranted to ensure timely diagnosis and appropriate management.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study protocol was approved by the Scientific Researches Evaluation and Ethics Committee No. 1 at Ankara Etlik City Hospital. (Date: 14.05.2025, Decision No: AESH-BADEK1-2025-204).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Ando H. Embryology of the biliary tract. *Dig Surg.* 2010;27(2):87-89. doi:10.1159/000286463
2. Dachman AH, Schneck C. Embryology of the gallbladder. In: imaging atlas of the normal gallbladder and its variants. Boca Raton, FL: CRC Press. 2018.
3. Ochshorn Y, Rosner G, Barel D, Bronshtein M, Muller F, Yaron Y. Clinical evaluation of isolated nonvisualized fetal gallbladder. *Prenat Diagn*. 2007;27(8):699-703. doi:10.1002/pd.1757
4. Zeng Y, Hu R, Lu J, et al. Prenatal genetic detection in foetus with gallbladder size anomalies: cohort study and systematic review of the literature. *Ann Med.* 2025;57(1):2440638. doi:10.1080/07853890.2024.2440638
5. Bardin R, Ashwal E, Davidov B, Danon D, Shohat M, Meizner I. Nonvisualization of the fetal gallbladder: can levels of γ-glutamyl transpeptidase in amniotic fluid predict fetal prognosis? *Fetal Diagn Ther.* 2016;39(1):50-55. doi:10.1159/000430440
6. Zhang H, Zhu X, Kang J, Sun Y, Yang H. Pregnancy outcomes of non-visualization of the fetal gallbladder from a Chinese tertiary single centre and literature review. *Children (Basel).* 2022;9(9):1288. doi:10.3390/children9091288
7. Avni FE, Garel C, Naccarella N, Franchi-Abella S. Anomalies of the fetal gallbladder: pre- and postnatal correlations. *Pediatr Radiol.* 2023;53(4):602-609. doi:10.1007/s00247-022-05457-w
8. Markova D, Markova T, Pandya P, David AL. Postnatal outcome after ultrasound findings of an abnormal fetal gallbladder: a systematic review and meta-analysis. *Prenat Diagn.* 2025;45(2):185-195. doi:10.1002/pd.6719
9. Karataş E, Tanaçan A, Özkavak OO, et al. Outcomes of pregnancies diagnosed with absent or abnormal fetal gallbladder in a tertiary center. *Int J Gynaecol Obstet.* 2025;168(3):1031-1038. doi:10.1002/ijgo.15949
10. Vij M, Rela M. Biliary atresia: pathology, etiology and pathogenesis. *Future Sci OA*. 2020;6(5):FSO466. doi:10.2144/fsoa-2019-0153
11. Tam PK, Wells RG, Tang CS, et al. Biliary atresia. *Nat Rev Dis Primers.* 2024;10(1):47. doi:10.1038/s41572-024-00533-x
12. Xu W, Ling W, Ren X, et al. Prenatal ultrasound features of biliary atresia: diagnostic significance of abnormal gallbladder size and hepatic hilar cyst. *Prenat Diagn.* 2025;45(2):185-195. doi:10.1002/pd.6865
13. Choi SO, Park WH, Lee HJ, Woo SK. Triangular cord: a sonographic finding applicable in the diagnosis of biliary atresia. *J Pediatr Surg.* 1996;31(3):363-366. doi:10.1016/s0022-3468(96)90739-3
14. Yoon HM, Suh CH, Kim JR, Lee JS, Jung AY, Cho YA. Diagnostic performance of sonographic features in patients with biliary atresia: a systematic review and meta-analysis. *J Ultrasound Med.* 2017;36(10):2027-2038. doi:10.1002/jum.14234
15. Napolitano M, Franchi-Abella S, Damasio MB, et al. Practical approach to imaging diagnosis of biliary atresia, part 1: prenatal ultrasound and magnetic resonance imaging, and postnatal ultrasound. *Pediatr Radiol.* 2021;51(2):314-331. doi:10.1007/s00247-020-04840-9
16. He M, Xie H, Du L, Lei T, Zhang L. Postnatal outcomes of fetuses with isolated gallbladder anomalies: be aware of biliary atresia. *J Matern Fetal Neonatal Med.* 2022;35(25):7005-7010. doi:10.1080/14767058.2021.1933936
17. Koob M, Pariente D, Habes D, et al. The porta hepatic microcyst: an additional sonographic sign for the diagnosis of biliary atresia. *Eur Radiol.* 2017;27(5):1812-1817. doi:10.1007/s00330-016-4546-5

18. Duguépéroux I, Scotet V, Audrézet MP, et al. Nonvisualization of fetal gallbladder increases the risk of cystic fibrosis. *Prenat Diagn*. 2012;32(1):21-28. doi:10.1002/pd.2866

19. Bergougnoux A, Jouannic JM, Verneau F, et al. Isolated nonvisualization of the fetal gallbladder should be considered for the prenatal diagnosis of cystic fibrosis. *Fetal Diagn Ther.* 2019;45(5):312-316. doi:10.1159/000489120

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1739961

Anatolian Curr Med J. 2025;7(5):647-651

The effect of vitamin D replacement therapy on insulin resistance: a retrospective study

Bahar Arıcan Tarım[1], Tuğçe Uskur[2]

[1]*Department of Internal Medicine, Kartal Dr. Lütfi Kırdar City Hospital, İstanbul, Turkiye*

[2]*Department of Medical Pharmacology, Faculty of Medicine, Kırklareli University, Kırklareli, Turkiye*

Cite this article as: Arıcan Tarım B, Uskur T. The effect of vitamin D replacement therapy on insulin resistance: a retrospective study. *Anatolian Curr Med J.* 2025;7(5):647-651.

Received: 11.07.2025 ◆ **Accepted**: 03.09.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: This study retrospectively evaluated the effect of a 3-month vitamin D replacement therapy in patients diagnosed with insulin resistance (IR) and low vitamin D levels. The relationship between changes in vitamin D levels and IR was investigated.

Methods: A total of 98 patients with low serum vitamin D levels and HOMA-IR >2.5, who were admitted to the internal medicine outpatient clinic between January 1 and December 31, 2023, were included in this retrospective study. Sociodemographic characteristics and pre-and post-treatment biochemical parameters were obtained from medical records. Post-supplementation levels of vitamin D, glucose, insulin, and HOMA-IR were analyzed. Pre-and post-treatment values were compared using paired t-tests. The relationship between changes in vitamin D levels and HOMA-IR scores was assessed using Spearman correlation analysis.

Results: Of the 98 patients, 85.7% were female and 14.3% were male, with a mean age of 39.09±11.45 years. A significant reduction in HOMA-IR score was observed after vitamin D supplementation (pre: 4.54±2.05; post: 3.01±1.28; $p<0.0001$). A statistically significant negative correlation was found between the increase in vitamin D levels and the decrease in HOMA-IR scores ($r=-0.298$, $p=0.0029$). In addition, glucose and insulin levels significantly decreased after treatment, while no significant changes were observed in lipid profile or vitamin B12 levels.

Conclusion: The findings suggest that vitamin D replacement therapy may have a beneficial effect on IR. Vitamin D supplementation could be considered as a supportive approach in the management of IR, especially in individuals with vitamin D deficiency. The more pronounced improvement observed in female patients indicates the potential importance of considering sex-specific differences in future clinical assessments, but this should be interpreted cautiously due to the limited number of male participants.

Keywords: Vitamin D, insulin resistance, HOMA-IR, diabetes

INTRODUCTION

Insulin resistance (IR) is clinically characterized by impaired glucose uptake and utilization due to inadequate cellular response to endogenous or exogenous insulin.[1] Recognized as a precursor to type 2 diabetes, this pathological response in glucose metabolism has emerged as a growing global public health concern.[2] The reported prevalence of IR ranges from 15.5% to 46.5% in the literature,[3,4] and it has been associated with all-cause mortality at rates of 20.6–25.3%.[5]

While IR is a component of metabolic syndrome, both genetic and environmental factors contribute to its etiology. Risk factors for IR include obesity, physical inactivity, aging, smoking, sedentary lifestyle, family history of diabetes, certain health conditions such as polycystic ovary syndrome (PCOS), and medications such as thiazide diuretics, β-adrenergic blockers, and glucocorticoids.[6] As diminished insulin action has been shown to play a key role in the pathophysiology of many diseases, identifying direct or indirect modulators of IR has gained significant importance. Accordingly, various metabolic conditions associated with IR have been widely studied.[7] One such factor is vitamin D, which has long been recognized for its influence on insulin action and its association with IR.[8]

Vitamin D is a fat-soluble secosteroid hormone primarily involved in the regulation of calcium and phosphorus metabolism and bone mineralization.[9] Beyond its classical roles, vitamin D has been shown to contribute to the prevention of various cancers, obesity, autoimmune, cardiovascular, and infectious diseases.[10]

In recent years, numerous studies have highlighted the relationship between vitamin D and IR. The notion that vitamin D deficiency may contribute to the development of

Corresponding Author: Tuğçe Uskur, tugceuskur@gmail.com

IR and/or worsen the course of metabolic disorders related to IR has gained strong support. Although the mechanisms underlying the development of IR in the context of vitamin D deficiency remain unclear, several studies have suggested that the anti-inflammatory and antioxidant properties of vitamin D may be responsible. Deficiency may increase oxidative stress and promote a pro-inflammatory state in hepatocytes, contributing to IR.[11] Another study indicated that vitamin D deficiency impairs glucose-stimulated insulin secretion by downregulating the expression of peroxisome proliferator-activated receptor gamma (PPAR-γ).[12] In addition to its preventive role against IR and type 2 diabetes mellitus (T2DM), vitamin D supplementation has been associated with improved glycemic control in diabetic patients. Given that vitamin D deficiency is frequently encountered in outpatient settings, timely diagnosis and treatment have become increasingly important.[13,14] The aim of this study is to retrospectively evaluate the effect of a 3-month vitamin D replacement therapy on IR in patients with low serum vitamin D levels.

METHODS

Ethics

This study was conducted as a retrospective analysis. Ethical approval was obtained from the Clinical Researches Ethics Committee of Kartal Dr. Lütfi Kırdar City Hospital (Date: 27.09.2023, Decision No: 2023/514/258/1). The study was carried out at Kartal Dr. Lütfi Kırdar City Hospital in accordance with the principles of the Declaration of Helsinki.

Study Population

The study was conducted with patients who were diagnosed with IR and vitamin D deficiency and met the inclusion criteria at the internal medicine outpatient clinic of Kartal Dr. Lütfi Kırdar City Hospital between January 1, 2023 and December 31, 2023.

Inclusion Criteria

- Patients diagnosed with IR at the internal medicine outpatient clinic between 01/01/2023 and 31/12/2023, followed up by a single clinician
- Patients not using any pharmacological treatment for IR
- Aged between 18 and 65 years
- No diagnosis of diabetes mellitus
- Serum 25-hydroxy vitamin D (25-OH-D) levels <20 ng/ml
- Received vitamin D replacement therapy for 3 months following the diagnosis of vitamin D deficiency, applied under the same treatment algorithm and supervised by the clinician

Exclusion Criteria

- Patients using medications for IR (e.g., metformin, pioglitazone)
- Diagnosed with diabetes mellitus
- Serum 25-OH-D vitamin levels >20 ng/ml
- Age >65 or <18 years
- Patients with terminal-stage malignancies
- Patients with advanced neurological or psychiatric disorders
- Pregnant women

Parameters Evaluated at Baseline and After 3-Month Vitamin D Supplementation

- Age
- Vitamin D
- Glucose
- Insulin
- Vitamin B12
- HbA1c
- HDL cholesterol
- LDL cholesterol
- Triglycerides
- Total cholesterol

Calculation of HOMA-IR

The homeostatic model assessment (HOMA) is a mathematical model used to evaluate IR based on fasting glucose and insulin concentrations.[15] In this study, IR was defined as a HOMA-IR score greater than 2.5, calculated using the following formula:

HOMA-IR score=(fasting glucose×gasting insulin)/405

Statistical Analysis

The data analyses of the patient data collected in this study were performed using GraphPad Prism version 8.0 software. Continuous variables were evaluated using descriptive statistics and expressed as mean±standard deviation (mean±SD). Paired t-test or Wilcoxon signed-rank test was used to compare clinical and biochemical parameters before and after vitamin D supplementation. Paired t-test was specifically used for evaluating changes in HOMA-IR scores. To assess the relationship between the change in vitamin D levels and the change in HOMA-IR scores, Spearman correlation analysis was conducted. A p-value less than 0.05 was considered statistically significant.

RESULTS

Among the 98 patients included in the study, 14.3% (n=14) were male and 85.7% (n=84) were female. The mean age of the patients was 39.09±11.45 years (range: 20–65), with males having a mean age of 34.57±12.4 years (range: 20–65) and females 38.03±11.21 years (range: 20–60).

At the time of the initial visit, the patients' clinical and biochemical parameters were as follows: mean serum vitamin D level was 11±3.94 ng/ml, vitamin B12 level was 280±129.6 pg/ml, glucose level was 94.17±9.82 mg/dl, HbA1c level was 5.54±0.34%, insulin level was 19.52±8.27 µIU/ml, HDL

cholesterol was 50.59±10.83 mg/dl, LDL cholesterol was 114.36±30.64 mg/dl, triglycerides were 138.69±71.5 mg/dl, and total cholesterol was 193.37±38.57 mg/dl (**Table 1**).

Table 1. Mean±standard deviation values of patients' clinical and biochemical parameters before vitamin D supplementation

Parameters	Total (n=98)	Male (n=14)	Female (n=84)
Year	39.09±11.45	34.57±12.4	38.03±11.21
Vitamin D (ng/ml)	11±3.94	12.97±3.55	10.67±3.92
Glucose (mg/dl)	94.17±9.82	95.71±8.07	93.91±10.1
Insulin (µIU/ml)	19.52±8.27	17.2±5.25	19.91±8.63
Vitamin B12 (pg/ml)	280±129.6	287.78±85.87	278.7±135.88
HbA1c (%)	5.54±0.34	5.61±0.23	5.53±0.35
HDL (mg/dl)	50.59±10.83	44.21±7.47	51.65±10.97
LDL (mg/dl)	114.36±30.64	119.5±19.8	113.49±32.12
Triglycerides (mg/dl)	138.69±71.5	136.71±49.71	139.02±74.74
Total cholesterol (mg/dl)	193.37±38.57	190.92±22.43	193.78±40.73

HDL: High-density lipoprotein, LDL: Low density lipoprotein

The mean HOMA-IR score, calculated using fasting glucose and insulin levels of the 98 patients included in the study, was 4.54±2.05 at the initial visit. Following vitamin D supplementation, the mean HOMA-IR score was found to be 3.01±1.28 based on the blood samples obtained at follow-up visits. When the same analysis was performed based on gender, a significant reduction was observed in both sexes; however, this reduction was more pronounced in female patients ([Female HOMA-IR score before: 4.62±2.15, after: 2.97±1.35, p<0.0001]; [Male HOMA-IR score before: 4.05±1.19, after: 3.21±0.83, p=0.0011]). The comparison of HOMA-IR scores before and after vitamin D supplementation revealed a statistically significant decrease in HOMA-IR levels (p<0.0001) (**Figure 1**). The mean difference was–1.53, which was statistically significant (t(97)=11.87, p<0.0001, 95% CI:–1.789 to –1.277). The pairing was strong and significant (r=0.80).

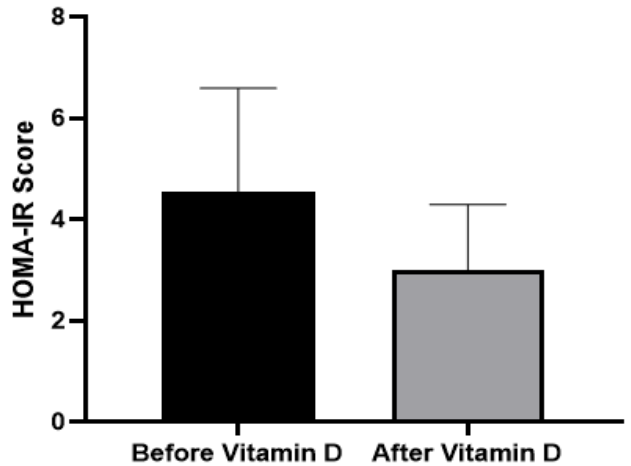


Figure 1. Comparison of HOMA-IR scores before and after vitamin D supplementation (n=98; mean difference=–1.53; p<0.0001)
HOMA-IR: Homeostatic model assessment-insulin resistance

Additionally, the proportion of patients whose HOMA-IR score dropped below 2.5 following vitamin D supplementation was also evaluated. Among the 98 patients with pre-treatment HOMA-IR ≥2.5, post-treatment values fell below 2.5 in 39 patients (39.8%), which was found to be statistically significant (p<0.0001). In the gender-based analysis, this reduction remained statistically significant in females (p<0.0001); however, due to the limited number of male patients whose HOMA-IR score dropped below 2.5, the difference was not statistically significant in males (p=1.00) (**Table 2**).

Table 2. Comparison of HOMA-IR scores before and after vitamin D supplementation by gender

Group	HOMA-IR before (mean±SD)	HOMA-IR after (mean±SD)	p-value	Patients with HOMA-IR <2.5 (n, %)	p-value (Drop <2.5)
Total	4.54±2.05	3.01±1.28	p<0.0001	39 (39.8%)	p<0.0001
Female	4.62±2.15	2.97±1.35	p<0.0001	38 (45.2%)	p<0.0001
Male	4.05±1.19	3.21±0.83	p=0.0011	1 (6.6%)	p=1.00

Vitamin D replacement therapy the rate of patients with a HOMA-IR score below 2.5 was found to be 39.8% overall, 6.6% in men, and 45.2% in women. HOMA-IR: Homeostatic model assessment-insulin resistance, SD: Standard deviation

A statistically significant negative correlation was found between the change in vitamin D levels and the change in HOMA-IR scores in patients (Spearman r=–0.298, p=0.0029) (**Figure 2**).

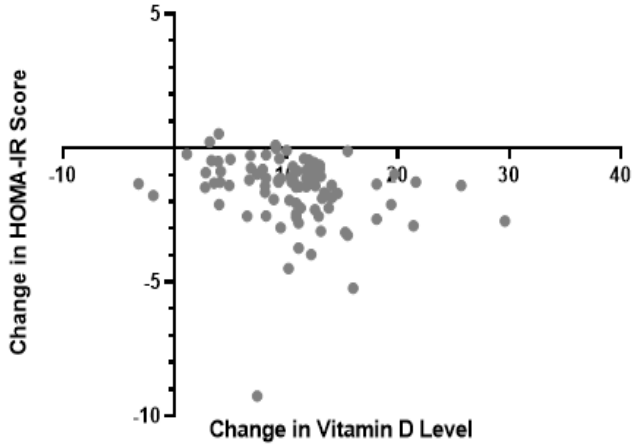


Figure 2. Correlation between the change in vitamin D levels and the change in HOMA-IR scores (Spearman r=–0.298, p=0.0029)
HOMA-IR: Homeostatic model assessment-insulin resistance

Following 3 months of vitamin D supplementation, the patients' mean vitamin D level increased significantly, indicating the effectiveness of the treatment (pre-treatment: 11±3.94 ng/ml, post-treatment: 21.68±6.18 ng/ml; p<0.0001). A significant decrease was also observed in glucose and insulin levels after supplementation. However, no statistically significant changes were found in B12, HbA1c, HDL, LDL, triglycerides, or total cholesterol levels (**Table 3**).

DISCUSSION

IR is a reduced biological response to exogenous or endogenous insulin and a major public health problem linked to multiple comorbidities.[2] It is influenced by genetics, perinatal factors, sex, ethnicity, puberty, medications, obesity, lifestyle and nutrition.[6,16,17] One of the factors affecting IR, vitamin D has been widely studied for its potential role in glucose metabolism. It is known to enhance insulin production and insulin

Table 3. Comparison of biochemical parameters before and after vitamin D supplementation

Parameters	Before (mean±SD)	After (mean±SD)	p-value
Vitamin D	11±3.94	21.68±6.18	<0.0001
Glucose	94.17±9.82	91.42±8.68	0.0001
Insulin	19.52±8.27	13.28±5.30	<0.0001
Vitamin B12	280±129.60	275.03±120.15	0.7442
HbA1c	5.54±0.34	5.55±0.31	0.6427
HDL	50.59±10.83	51.14±10.48	0.3484
LDL	114.36±30.64	116.9±27.33	0.3121
Triglycerides	138.69±71.5	137.78±55.69	0.8290
Total cholesterol	193.37±38.57	195.74±34.63	0.2552

SD: Standard deviation, HDL: High-density lipoprotein, LDL: Low density lipoprotein

receptor expression, while also reducing proinflammatory cytokine levels.[18] The important roles of vitamin D in glucose metabolism are known, and the pathways involved are still under investigation.[8,19]

In this study, which aimed to investigate IR and various clinical and biochemical parameters before and after vitamin D treatment in patients diagnosed with IR, 14.3% (n=14) of the 98 included patients were male and 85.7% (n=84) were female. The high prevalence of IR among female patients is consistent with literature data showing that women tend to have more than men.[20] This can be explained by certain biological disadvantages in women that play a role in the pathophysiology of IR, including gender hormones, environmental and lifestyle factors, differences in body fat and muscle mass distribution and function, and genetic variations, particularly prominent during menopause.[21] Among other sociodemographic data examined in our study, the mean age was 39.09±11.45 years (range 20–65), with 34.57±12.4 in males (range 20–65) and 38.03±11.21 in females (range 20–60). This is consistent with the age data reported by Özdin et al.[22] in 2021.

In line with our study aim, the mean HOMA-IR score was calculated using fasting glucose and insulin levels of the 98 patients at initial admission was 4.54±2.05. After 3 months of vitamin D supplementation, the mean HOMA-IR score was found to be 3.01±1.28, and the difference between the pre- and post-treatment scores was statistically significant (p<0.0001). A statistically significant negative correlation was found between the change in vitamin D levels and the change in HOMA-IR scores. In a study conducted in Japan, a comparison between a group receiving vitamin D supplementation and a placebo group showed a significant decrease in HOMA-IR levels in the supplementation group after one year.[23] Another noteworthy study reported a 60% improvement in insulin sensitivity after one year of vitamin D treatment, suggesting that this treatment may be more effective than troglitazone or metformin.[24] In a 4-year randomized controlled trial at the Medical University of Graz, Austria, 150 healthy premenopausal women with 25(OH)D levels <75 nmol/L received 20.000 IU vitamin D once weekly for 24 weeks. Importantly, this shorter-term intervention, similar in duration to our study, showed significant improvements in HOMA-IR, supporting our findings.[25]

In our study, it was found that only three months of vitamin D replacement therapy the rate of patients with a HOMA-IR score below 2.5 was found to be 39.8% overall, 6.6% in men, and 45.2% in women. It is noteworthy that vitamin D supplementation significantly reduced HOMA-IR scores more pronounced effect in women. Several studies suggest that vitamin D supplementation may exert a stronger protective effect against IR in women. Von Hurst et al.[26] reported improved insulin sensitivity following vitamin D supplementation in insulin-resistant women. Additionally, evidence from another study indicated that found sex-specific associations between vitamin D status and prediabetes, with lower vitamin D levels linked to increased risk in men but not in women.[27] A narrative review published in 2023 further highlights potential synergistic interactions between vitamin D and estrogen, which may enhance the hormone's efficacy in improving insulin metabolism in women.[28] The more pronounced reduction in HOMA-IR observed in women may be explained by potential synergistic interactions between vitamin D and estrogen, which enhance β-cell function and insulin sensitivity. Additionally, vitamin D may reduce proinflammatory cytokine levels, preserving insulin signaling pathways. These mechanisms may account for the sex-specific differences observed in our study.

Limitations

An important limitation of our study is its retrospective design, which precludes full control over variables such as vitamin D dosage. In addition, data regarding potential weight changes, dietary interventions, physical activity, lifestyle modifications and concomitant medication use during the 3-month follow-up period were not available. However, based on clinical follow-up notes, no obvious weight loss was noted in these patients. Notably, all patients included in this study were followed by a single clinician, and vitamin D replacement therapy was applied under a consistent treatment algorithm. This consistency enhances the reliability of the follow-up data and reduces variability that could arise from different clinical practices.

Moreover, the 3-month follow-up period might seem short, but in clinical practice, patients receiving vitamin D supplementation are typically re-evaluated after 3 months. In this context, the relatively short follow-up period can actually be advantageous, as it minimizes the potential influence of confounding factors such as lifestyle changes, weight variations, diet, and physical activity. Importantly, even within this 3-month period, a significant reduction in HOMA-IR was observed.

Nevertheless, the short duration still limits assessment of the long-term sustainability of the observed effects. Therefore, to clarify the causal relationship between vitamin D and IR, particularly in women, for whom this condition is a major public health concern and to elucidate the mechanisms underlying the more pronounced effectiveness of vitamin D in women, larger prospective and randomized controlled studies are needed.

CONCLUSION

As a result, this study clearly demonstrates that vitamin D therapy may have a beneficial effect on IR. Therefore, vitamin D supplementation could be considered as an adjunctive option in the management of patients diagnosed with IR. The more pronounced improvement observed in female patients should be interpreted cautiously and may indicate the potential importance of considering sex-specific differences in treatment planning, given the small number of male participants.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was initiated with the approval of the Kartal Dr. Lütfi Kırdar City Hospital Clinical Researches Ethics Committee (Date: 27.09.2023, Decision No: 2023/514/258/1).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Liu X, Tang HY, Luo ZC. Insulin resistance and skin diseases. *Zhongguo Yi Xue Ke Xue Yuan Xue Bao*. 2020;42(2):247-250. doi:10.3881/j.issn.1000-503X.11609
2. Goh LPW, Sani SA, Sabullah MK, Gansau JA. The prevalence of insulin resistance in Malaysia and Indonesia: an updated systematic review and meta-analysis. *Medicina (Kaunas)*. 2022;58(6):826. doi:10.3390/medicina58060826
3. Bermudez V, Salazar J, Martínez MS, et al. Prevalence and associated factors of insulin resistance in adults from Maracaibo City, Venezuela. *Adv Prev Med*. 2016;2016:9405105. doi:10.1155/2016/9405105
4. Qu HQ, Li Q, Rentfro AR, Fisher-Hoch SP, McCormick JB. The definition of insulin resistance using HOMA-IR for Americans of Mexican descent using machine learning. *PLoS One*. 2011;6(6):e21041. doi:10.1371/journal.pone.0021041
5. Penno G, Solini A, Orsi E, et al. Insulin resistance, diabetic kidney disease, and all-cause mortality in individuals with type 2 diabetes: a prospective cohort study. *BMC Med*. 2021;19(1):66. doi:10.1186/s12916-021-01936-3
6. Insulin Resistance & Prediabetes. National Institute of Diabetes and Digestive and Kidney Diseases. 2018.
7. Lazear J, Kapustin J. Vitamin D deficiency and type 2 diabetes: a retrospective review. *J Nurse Practitioners*. 2014;10(3):175-182. doi:10.1016/j.nurpra.2013.12.010
8. Chiu KC, Chu A, Go VL, Saad MF. Hypovitaminosis D is associated with insulin resistance and beta cell dysfunction. *Am J Clin Nutr*. 2004;79(5):820-825. doi:10.1093/ajcn/79.5.820
9. Trimarco V, Manzi MV, Mancusi C, et al. Insulin resistance and vitamin D deficiency: a link beyond the appearances. *Front Cardiovasc Med*. 2022;9:859793. doi:10.3389/fcvm.2022.859793
10. Yetgin S, Ozsoylu S, Ruacan S, Tekinalp G, Sarialioğlu F. Vitamin D-deficiency rickets and myelofibrosis. *J Pediatr*. 1989;114(2):213-217. doi:10.1016/s0022-3476(89)80785-1
11. Tao S, Yuan Q, Mao L, Chen FL, Ji F, Cui ZH. Vitamin D deficiency causes insulin resistance by provoking oxidative stress in hepatocytes. *Oncotarget*. 2017;8(40):67605-67613. doi:10.18632/oncotarget.18754
12. Park S, Kim DS, Kang S. Vitamin D deficiency impairs glucose-stimulated insulin secretion and increases insulin resistance by reducing PPAR-γ expression in nonobese type 2 diabetic rats. *J Nutr Biochem*. 2016;27:257-265. doi:10.1016/j.jnutbio.2015.09.013
13. Alissa EM, Alnahdi WA, Alama N, Ferns GA. Insulin resistance in Saudi postmenopausal women with and without metabolic syndrome and its association with vitamin D deficiency. *J Clin Transl Endocrinol*. 2014;2(1):42-47. doi:10.1016/j.jcte.2014.09.001
14. Lips P, Eekhoff M, van Schoor N, et al. Vitamin D and type 2 diabetes. *J Steroid Biochem Mol Biol*. 2017;173:280-285. doi:10.1016/j.jsbmb.2016.11.021
15. Wallace TM, Levy JC, Matthews DR. Use and abuse of HOMA modeling. *Diabetes Care*. 2004;27(6):1487-1495. doi:10.2337/diacare.27.6.1487
16. Gidding SS. Dyslipidemia in the metabolic syndrome in children. *J Cardiometab Syndr*. 2006;1(4):282-285. doi:10.1111/j.1559-4564.2006.05798.x
17. Chiarelli F, Marcovecchio ML. Insulin resistance and obesity in childhood. *Eur J Endocrinol*. 2008;159(Suppl 1):S67-74. doi:10.1530/EJE-08-0245
18. Mohan A, Haider R, Fakhor H, et al. Vitamin D and polycystic ovary syndrome (PCOS): a review. *Ann Med Surg*. 2023;85(7):3506-3511. doi:10.1097/MS9.0000000000000879
19. Ozfirat Z, Chowdhury TA. Vitamin D deficiency and type 2 diabetes. *Postgrad Med J*. 2010;86(1011):18-25. doi:10.1136/pgmj.2009.078626
20. Mittendorfer B. Insulin resistance: sex matters. *Curr Opin Clin Nutr Metab Care*. 2005;8(4):367-372. doi:10.1097/01.mco.0000172574.64019.98
21. Gado M, Tsaousidou E, Bornstein SR, Perakakis N. Sex-based differences in insulin resistance. *J Endocrinol*. 2024;261(1):e230245. doi:10.1530/JOE-23-0245
22. Özdin M, Yazar H, Mundan D. Sakarya Eğitim ve Araştırma Hastanesine başvuran hastaların homeostatik model değerlendirme-insülin direnci (HOMA-IR) değerlerinin yaş ve cinsiyet faktörü açısından değerlendirilmesi. *MKÜ Tıp Dergisi*. 2021;42(1):1-4. doi:10.17944/mkutfd.791428
23. Sun X, Cao ZB, Tanisawa K, Ito T, Oshima S, Higuchi M. Vitamin D supplementation reduces insulin resistance in Japanese adults: a secondary analysis of a double-blind, randomized, placebo-controlled trial. *Nutr Res*. 2016;36(10):1121-1129. doi:10.1016/j.nutres.2016.07.006
24. Chiu KC, Chu A, Go VL, Saad MF. Hypovitaminosis D is associated with insulin resistance and beta cell dysfunction. *Am J Clin Nutr*. 2004;79(5):820-825. doi:10.1093/ajcn/79.5.820
25. Trummer C, Theiler-Schwetz V, Kollmann M, et al. Effects of vitamin D supplementation on metabolic and endocrine parameters in healthy premenopausal women: a randomized controlled trial. *Clin Nutr*. 2020;39(3):718-726. doi:10.1016/j.clnu.2019.03.007
26. von Hurst PR, Stonehouse W, Coad J. Vitamin D supplementation reduces insulin resistance in South Asian women living in New Zealand who are insulin resistant and vitamin D deficient-a randomised, placebo-controlled trial. *Br J Nutr*. 2010;103(4):549-555. doi:10.1017/S0007114509992017
27. Ziyab AH, Mohammad A, Almousa Z, Mohammad T. Sex differences in the association between vitamin D and prediabetes in adults: a cross-sectional study. *Nutr Diabetes*. 2024;14(1):49. doi:10.1038/s41387-024-00311-4
28. Ciarambino T, Crispino P, Guarisco G, Giordano M. Gender differences in insulin resistance: new knowledge and perspectives. *Curr Issues Mol Biol*. 2023;45(10):7845-7861. doi:10.3390/cimb45100496

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1736977

Anatolian Curr Med J. 2025;7(5):652-656

Surgical outcomes of modified Ferguson hemorrhoidectomy for grade IV hemorrhoidal disease

Ahmet Cihangir Emral[1], Gökay Çetinkaya[2], Merter Gülen[1], Bahadır Ege[1]

[1]*Department of General Surgery, Medicana International Ankara Hospital, Faculty of Medicine, Atılım University, Ankara, Turkiye*

[2]*Department of General Surgery, Kartal Dr. Lütfi Kırdar City Hospital, University of Health Sciences, İstanbul, Turkiye*

Cite this article as: Emral AC, Çetinkaya G, Gülen M, Ege B. Surgical outcomes of modified Ferguson hemorrhoidectomy for grade IV hemorrhoidal disease. *Anatolian Curr Med J.* 2025;7(5):652-656.

Received: 07.07.2025 ◆ **Accepted**: 03.09.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: Grade IV internal hemorrhoidal disease represents an advanced stage of hemorrhoidal pathology, often necessitating surgical intervention. Ferguson hemorrhoidectomy, a closed excisional technique, is frequently employed for its reported advantages in postoperative recovery and complication profiles. This study aims to evaluate the effectiveness, postoperative outcomes, and complication rates of modified Ferguson hemorrhoidectomy in patients with grade IV internal hemorrhoidal disease.

Methods: In this retrospective analysis, 229 patients with grade IV internal hemorrhoidal disease who underwent modified Ferguson hemorrhoidectomy between March 2021 and December 2024 were included. Demographic data, postoperative complications, postoperative pain, recurrence rates, and time to return to daily activities were recorded. Postoperative follow-up was conducted up to a mean duration of 16.8 months.

Results: Among 229 patients (mean age 43.7±10.7 years; 58% male), 81% underwent three-quadrant hemorrhoidectomy. Postoperative complications included urinary retention (3.9%), urinary tract infection (0.9%), minor bleeding (1.3%), mild anal stenosis (0.9%), and minor perianal infection (5.2%). The median postoperative VAS pain scores were 4 (range: 2–7) on day 1 and 2 (range: 1–5) on day 7. No significant differences were observed in preoperative and postoperative Cleveland Clinic Incontinence Scores ($p>0.05$). Median time to return to normal daily activities was 12 days (range 5–19).

Conclusion: Modified Ferguson hemorrhoidectomy provides an effective and safe surgical option for grade IV internal hemorrhoidal disease, with low complication and recurrence rates and a favorable recovery profile. These results support its use as a reliable treatment approach in advanced hemorrhoidal disease.

Keywords: Ferguson hemorrhoidectomy, hemorrhoidal disease, postoperative complications, proctology, bleeding

INTRODUCTION

Hemorrhoidal disease (HD) is the most frequently encountered benign anorectal condition in proctologic practice. It is among the leading causes of lower gastrointestinal bleeding and typically presents clinically as painless rectal bleeding. In addition to bleeding, commonly observed symptoms include perianal swelling, prolapse, fecal soiling, pruritus ani, and anal discomfort.[1-3]

Various medical and surgical treatment options have been described for HD. Medical treatments include lifestyle modifications, dietary recommendations, and the use of topical or systemic medications. Surgical options for HD comprise Milligan-Morgan hemorrhoidectomy, Ferguson hemorrhoidectomy, LigaSure hemorrhoidectomy (LigaSure™, Valleylab, Covidien), stapled hemorrhoidopexy (PPH), rubber band ligation, laser ablation, Doppler-guided hemorrhoidal artery ligation, sclerotherapy, infrared coagulation, and radiofrequency ablation (RFA).[4,5] Internal hemorrhoids are classified according to the Goligher classification, which is based on the presence and severity of prolapse. Among these treatment modalities, surgical excision—particularly hemorrhoidectomy—remains the cornerstone of therapy in complex or advanced cases, particularly in patients with grade III or IV hemorrhoids.[4,6] Although non-life-threatening complications such as urinary retention, urinary tract infection (UTI), and minor bleeding are the most commonly observed after hemorrhoidectomy, serious and potentially life-threatening complications such as rectal perforation, major bleeding, retroperitoneal and pelvic abscesses, and sepsis may also occur.[7]

The primary aim of this study is to comprehensively evaluate the effectiveness, postoperative outcomes, and complication rates of modified Ferguson hemorrhoidectomy in patients with grade IV internal HD. This study seeks to enhance our

Corresponding Author: Ahmet Cihangir Emral, cihangiremral@hotmail.com

understanding of the surgical outcomes associated with this advanced form of HD.

METHODS

This study received ethical approval from the Ethics Committee of Atılım University Faculty of Medicine, Medicana International Ankara Hospital (Date: 09.11.2023, Decision No: 32). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki. Between March 2021 and December 2024, a total of 2,534 patients presenting to our proctology clinic with HD were retrospectively evaluated. Among these, 769 patients underwent surgical intervention. Of these, 229 patients who underwent modified Ferguson hemorrhoidectomy for grade IV internal HD were included in the study.

Patients who underwent surgical procedures other than modified Ferguson hemorrhoidectomy; those diagnosed with grade I, II, or III internal HD; patients with uncontrolled diabetes, pregnancy, inflammatory bowel disease, inability to tolerate general anesthesia, a history of pelvic/perianal radiotherapy, bleeding disorders, or previous perianal surgery were excluded from the study.

All patients underwent a preoperative anorectal examination in the proctology unit. For patients over 50 years of age, colonoscopy was additionally performed. Based on the extent of the HD, each patient was treated with modified Ferguson hemorrhoidectomy involving two or three quadrants. All surgeries were performed by Dr. B.E., an experienced proctologist.

All patients were discharged on the first postoperative day. Preoperatively, all patients received intravenous ciprofloxacin 500 mg and metronidazole 500 mg as antibiotic prophylaxis. All patients received standardized postoperative care. Postoperatively, oral antibiotic therapy consisting of ciprofloxacin 500 mg twice daily and metronidazole 500 mg three times daily was prescribed for 5 days, along with micronized flavonoid fraction 1000 mg twice daily for one month. The routine analgesic regimen consisted of oral dexketoprofen 25 mg twice daily and diclofenac sodium SR 100 mg once daily. Intramuscular NSAIDs were not routinely prescribed; however, they were administered in cases where oral analgesics were insufficient. In addition, all patients were advised to take warm sitz baths twice daily, use stool softeners, and follow a high-fiber diet with adequate hydration.

Patients were invited for routine follow-up visits on postoperative day 1, at week 1, week 2, month 1, month 6, and year 1. Following the first postoperative year, patients were advised to attend annual proctology follow-up visits. Patients were advised to return for evaluation in case of any ongoing symptoms. Demographic characteristics, postoperative complications, postoperative pain assessed by the Visual Analog Scale (VAS) on postoperative days 1 and 7, time to return to normal daily activities, and recurrence rates were recorded using standardized, prospectively completed forms.

Surgical Procedure

All patients underwent surgery under general anesthesia combined with local infiltration of bupivacaine. With the patient in the prone jackknife position, an elliptical incision was made over the diseased hemorrhoidal tissue. Dissection was performed in the submucosal plane, carefully preserving the internal sphincter muscle, while intervening skin bridges were left intact to avoid circumferential scarring. Hemorrhoidal tissue was excised up to the anorectal junction using electrocautery, taking care to preserve intervening skin bridges.

The pedicle was ligated with a continuous 3/0 polyglactin suture, which was also used to close the rectal mucosa. Hemostasis was achieved by a combination of electrocautery, suture ligation, and the placement of an Absorbalb Anal Sponge (8×3 cm; SPONGOSTAN™, Ethicon) in the anal canal at the end of the procedure. No sutures were placed in the anoderm, a deliberate modification intended to reduce postoperative pain, avoid ischemia, and minimize the risk of anal stenosis. This modification in the technique represents the key difference in the procedure. Postoperative dressing consisted solely of the anal sponge. The surgical procedure is illustrated in **Figure**.

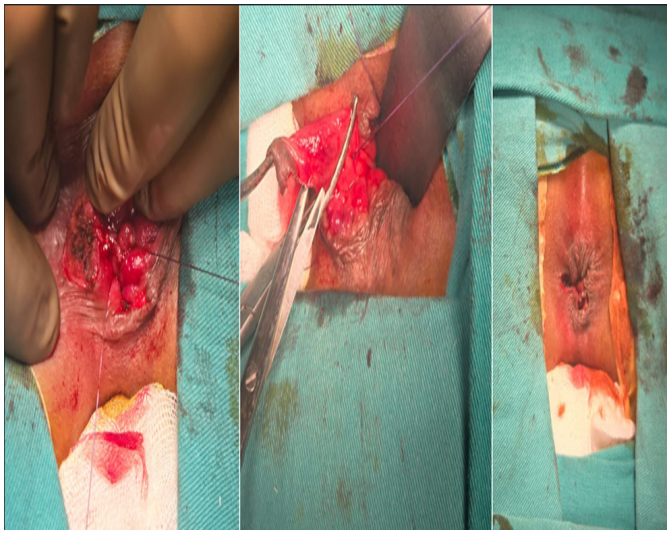


Figure. Intraoperative view of a modified Ferguson hemorrhoidectomy for grade IV hemorrhoids, performed without suturing the anoderm

Statistical Analysis

Data analysis was performed using SPSS version 27 (IBM Corp., USA). Normality of continuous variables was assessed using the Skewness and Kurtosis values. Variables with normal distribution were expressed as mean±SD, whereas non-normally distributed variables were summarized as median (IQR). Categorical data were presented as counts and percentages with 95% confidence intervals (CIs). Comparisons between preoperative and postoperative Cleveland Clinic Incontinence Scores (CCIS), which were not normally distributed due to skewness, were performed using the Wilcoxon signed-rank test. A p-value<0.05 was considered statistically significant.

RESULTS

A total of 229 patients, including 133 males and 96 females, were included in the study. The demographic characteristics of the patients and details of the quadrant-based hemorrhoidectomy are presented in **Table 1**.

Table 1. Demographic characteristics and extent of hemorrhoidectomy

Characteristic	n (%) or mean±SD
Age (years)	43.7±10.7
Sex	
Female	96 (42)
Male	133 (58)
Hemorrhoidectomy (quadrants)	
Two quadrants	44 (19)
Three quadrants	185 (81)

SD: Standard deviation

All patients were discharged on postoperative day 1. During postoperative follow-up, complications included urinary retention in 9 patients (3.9%) and UTI in 2 patients (0.9%). Among those with urinary retention, 2 were female and 7 were male, with a mean age of 65.2 years. Both patients who developed UTIs were female, aged 64 and 74 years, respectively.

Postoperative bleeding was observed in 3 patients (1.3%). One of these was a 56-year-old male without a known anticoagulant history but with a recent history of using an unspecified herbal medication with anticoagulant properties. He was evaluated under general anesthesia; minor oozing was observed and bleeding was controlled surgically. The remaining two patients were managed conservatively with one day of observation and showed no further bleeding.

Anal stenosis, classified as mild according to Milsom and Mazier criteria, developed in 2 patients (0.9%). Both were male, aged 46 and 50 years, and were successfully treated with anal dilatation and internal sphincterotomy. None of the patients experienced postoperative deterioration in continence. The median CCIS values were 0 (IQR 0–0) both preoperatively and postoperatively, with no statistically significant difference (Wilcoxon signed-rank test, p>0.05).

Minor perianal infection was observed in 12 patients (5.2%) and was controlled with simple wound care. No major perianal infections or cases of Fournier's gangrene were detected. The median time to return to normal daily activities was 12 days (IQR 7-12).

The median postoperative VAS pain score was 4 (IQR 4–4) on day 1 and 2 (IQR 1–2) on day 7. Standard analgesia with oral NSAIDs was sufficient in the majority of patients, while only seven patients (3.0%) required additional intramuscular diclofenac injections for breakthrough pain.

No recurrence was observed in patients who underwent three-quadrant hemorrhoidectomy. Among the patients who underwent two-quadrant hemorrhoidectomy, recurrence of HD from the non-operated quadrant was observed in 5 patients (2.1%) within 1 to 2 years postoperatively. These patients were treated with rubber band ligation.

The mean follow-up period was 16.8±7.9 months. Postoperative outcomes are summarized in **Table 2**.

Table 2. Postoperative outcomes of patients

Outcome	Value (n=229)
Urinary retention	9 (3.9%, 95% CI: 2.1–7.3)
Urinary tract infection	2 (0.9%, 95% CI: 0.2–3.1)
Postoperative bleeding	3 (1.3%, 95% CI: 0.4–3.8)
Anal stenosis (Milsom and Mazier)	
Mild	2 (0.9%, 95% CI: 0.2–3.1)
Moderate	0
Severe	0
Perianal infection	
Minör	12 (5.2%, 95% CI: 3.0–8.9)
Major/Fournier's gangrene	0
Cleveland Clinic Inkontinence Score	Preop: 0 (IQR 0–0), Postop: 0 (IQR 0–0), p > 0.05 (Wilcoxon)
Postoperative pain (VAS)	Day 1: 4 (IQR 4–4), day 7: 2 (IQR 1–2)
Return to normal life, days	12 (IQR 7–12)
Recurrence	5 (2.1%, 95% CI: 0.9–4.9)
Follow-up, months	16.8±7.9

CI: Confidence interval, IQR: Interquartile range, VAS: Visual Analog Scale

DISCUSSION

Excisional surgical methods remain the most effective treatment option for advanced-stage HD, particularly grade 4, in terms of recurrence prevention. Among excisional hemorrhoidectomy techniques, the Ferguson (closed) hemorrhoidectomy has been shown to result in less postoperative pain, faster recovery, and lower rates of postoperative bleeding compared to the Milligan-Morgan (open) technique.[3,8] In our surgical practice, we routinely prefer the modified Ferguson hemorrhoidectomy technique.

Postoperative urinary retention has been reported in rates as high as 13%, with male gender and advanced age identified as significant risk factors. In this study, urinary retention was more frequent among older male patients; however, the incidence was lower than that reported in the literature. We believe this lower incidence is because all our surgeries were performed under general anesthesia. This is important since spinal anesthesia has been shown to increase the risk of postoperative urinary retention.[9,10] Bleday et al.[11] reported a UTI rate of 3.3% in their study. In contrast, the incidence of UTI in our cohort was markedly lower. We attribute this difference primarily to the use of general anesthesia, as previously noted, which may have contributed to a reduced incidence of postoperative urinary retention and, consequently, a lower risk of UTI. In addition, patients who received general anesthesia were mobilized earlier compared to those who would have received spinal anesthesia, and they were not affected by side effects such as spinal anesthesia–related postoperative headaches.

Postoperative bleeding following Ferguson hemorrhoidectomy has been described in the literature at rates up to 3%. It has been reported that males are more prone to postoperative bleeding, and that LigaSure hemorrhoidectomy carries a higher bleeding risk compared to Ferguson hemorrhoidectomy. Consistent with the literature, our postoperative bleeding rate was low. The use of continuous sutures in both the Ferguson and modified Ferguson hemorrhoidectomy techniques

contributes to the low incidence of postoperative bleeding. Moreover, Kaidar-Person et al.[12] have demonstrated that surgical experience plays a critical role in reducing bleeding complications.[3,12-14]

The time to return to normal activities in this study was comparable to that reported in the literature. Although fecal incontinence rates up to 6% and anal stenosis rates up to 3.8% have been reported after hemorrhoidectomy, this study found no significant difference between preoperative and postoperative CCIS, and only 0.9% of patients developed mild anal stenosis. We believe this low complication rate is due to our surgical expertise and specialization in proctology.[3,13,15,16]

Postoperative pain is the most significant complaint among patients following hemorrhoidectomy. Although pain after Ferguson (closed) hemorrhoidectomy is less than after Milligan-Morgan (open) hemorrhoidectomy, studies have shown that pain following Ligasure hemorrhoidectomy is even lower.[17,18] In their study, Khanna et al.[18] reported a mean VAS score of 5.2 on postoperative day 1 following Ferguson hemorrhoidectomy. In this study, although postoperative VAS scores were moderately low, it should be noted that all patients had advanced grade 4 disease, necessitating extensive resections.

Our modification can be regarded as an anoderm-sparing variation of the Ferguson hemorrhoidectomy. Unlike the classical Ferguson technique, in which the rectal mucosa, anoderm, and perianal skin are closed continuously, our method deliberately avoids anoderm suturing. This approach is conceptually similar to previously described anoderm-sparing techniques, with the rationale of reducing postoperative pain and the risk of anal stenosis by minimizing ischemia and tension in the anoderm. Although our findings showed relatively low pain scores, these observations should be interpreted with caution due to the retrospective, single-arm design of the study. Prospective randomized studies are needed to confirm whether this modification provides a true advantage over other closed or anoderm-sparing techniques.

Recurrence rates after Ferguson hemorrhoidectomy have been reported around 3%. In this study, no recurrence was observed in patients undergoing three-quadrant hemorrhoidectomy. However, among patients who underwent two-quadrant hemorrhoidectomy, HD developed in untreated quadrants during follow-up. We consider these cases as new disease developments rather than true recurrences.[12,13]

Limitations

The most significant limitation of this study is its retrospective design and the lack of a control/comparison group. In addition, routine 5-day antibiotic prophylaxis was employed as part of our institutional practice, which does not reflect current international guidelines recommending single-dose preoperative prophylaxis. This discrepancy should be considered when interpreting our results.

CONCLUSION

The modified Ferguson hemorrhoidectomy method offers an effective treatment for grade 4 internal HD, characterized by low complication rates and high treatment success. It is believed that postoperative pain is reduced in this method, particularly due to the absence of sutures placed in the anoderm. These results indicate that this method is a safe and reliable option for managing this advanced stage of the disease.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was initiated with the approval of the Atılım University Faculty of Medicine-affiliated Medicana International Ankara Hospital Ethics Committee (Date: 09.11.2023, Decision No: 32).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Lohsiriwat V, Jitmungngan R. Strategies to reduce post-hemorrhoidectomy pain: a systematic review. *Medicina (Kaunas).* 2022; 58(3): 418. doi:10.3390/medicina58030418
2. Gallo G, Sacco R, Sammarco G. Epidemiology of hemorrhoidal disease. In: Ratto C, Parello A, Litta F. (eds) Hemorrhoids. Coloproctology, vol 2. Springer, Cham. 2018. doi:10.1007/978-3-319-53357-5_1
3. Gallo G, Martellucci J, Sturiale A, et al. Consensus statement of the Italian society of colorectal surgery (SICCR): management and treatment of hemorrhoidal disease. *Tech Coloproctol.* 2020 Feb;24(2):145-164. doi:10.1007/s10151-020-02149-1
4. Simillis C, Thoukididou SN, Slesser AA, Rasheed S, Tan E, Tekkis PP. Systematic review and network meta-analysis comparing clinical outcomes and effectiveness of surgical treatments for haemorrhoids. *Br J Surg.* 2015;102(13):1603-1618. doi:10.1002/bjs.9913
5. Gülen M, Emral AC, Ege B. Hemoroid lastik band ligasyonu komplikasyonlarının yönetimi: masif rektal kanama. *Namik Kemal Med J.* 2024;12(1):40-44. doi:10.4274/nkmj.galenos.2023.36036
6. Sobrado Júnior CW, Obregon CA, E Sousa Júnior AHDS, Sobrado LF, Nahas SC, Cecconello I. A new classification for hemorrhoidal disease: the creation of the "BPRST" staging and its application in clinical practice. *Ann Coloproctol.* 2020;36(4):249-255. doi:10.3393/ac.2020.02.06
7. Klang E, Sobeh T, Amitai MM, Apter S, Barash Y, Tau N. Post hemorrhoidectomy complications: CT imaging findings. *Clin Imaging.* 2020;60(2):216-221. doi:10.1016/j.clinimag.2019.12.015
8. Bhatti MI, Sajid MS, Baig MK. Milligan-Morgan (open) versus Ferguson haemorrhoidectomy (closed): a systematic review and meta-analysis of published randomized, controlled trials. *World J Surg.* 2016;40(6):1509-1519. doi:10.1007/s00268-016-3419-z

9. Jeong HY, Song SG, Lee JK. Predictors of postoperative urinary retention after semiclosed hemorrhoidectomy. *Ann Coloproctol.* 2022;38(1):53-59. doi:10.3393/ac.2021.00304.0043
10. Tammela T. Postoperative urinary retention--why the patient cannot void. *Scand J Urol Nephrol Suppl.* 1995;175:75-77.
11. Bleday R, Pena JP, Rothenberger DA, Goldberg SM, Buls JG. Symptomatic hemorrhoids: current incidence and complications of operative therapy. *Dis Colon Rectum.* 1992;35(5):477-481. doi:10.1007/BF02049406
12. Kaidar-Person O, Person B, Wexner SD. Hemorrhoidal disease: a comprehensive review. *J Am Coll Surg.* 2007;204(1):102-117. doi:10.1016/j.jamcollsurg.2006.08.022
13. Milone M, Maietta P, Leongito M, Pesce G, Salvatore G, Milone F. Ferguson hemorrhoidectomy: is still the gold standard treatment? *Updates Surg.* 2012;64(3):191-194. doi:10.1007/s13304-012-0155-2
14. Lee KC, Liu CC, Hu WH, Lu CC, Lin SE, Chen HH. Risk of delayed bleeding after hemorrhoidectomy. *Int J Colorectal Dis.* 2019;34(2):247-253. doi:10.1007/s00384-018-3176-6
15. Guenin MO, Rosenthal R, Kern B, Peterli R, von Flüe M, Ackermann C. Ferguson hemorrhoidectomy: long-term results and patient satisfaction after Ferguson's hemorrhoidectomy. *Dis Colon Rectum*. 2005;48(8):1523-1527. doi:10.1007/s10350-005-0084-y
16. Eu KW, Teoh TA, Seow-Choen F, Goh HS. Anal stricture following haemorrhoidectomy: early diagnosis and treatment. *Aust N Z J Surg.* 1995;65(2):101-103. doi:10.1111/j.1445-2197.1995.tb07270.x
17. Lohsiriwat V, Jitmungngan R. Strategies to reduce post-hemorrhoidectomy pain: a systematic review. *Medicina (Kaunas).* 2022;58(3):418. doi:10.3390/medicina58030418
18. Khanna R, Khanna S, Bhadani S, Singh S, Khanna AK. Comparison of ligasure hemorrhoidectomy with conventional Ferguson's hemorrhoidectomy. *Indian J Surg.* 2010;72(4):294-297. doi:10.1007/s12262-010-0192-3

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1756949

Anatolian Curr Med J. 2025;7(5):657-664

Prognostic factors and treatment outcomes in chondrosarcoma: a single-institution experience

Tuğba Önder, Ülkü Yalçıntaş Arslan, Öztürk Ateş, Necati Alkış

Department of Medical Oncology, Dr. Abdurrahman Yurtaslan Ankara Oncology Training and Research Hospital, University of Health Sciences, Ankara, Turkiye

Cite this article as: Önder T, Yalçıntaş Arslan Ü, Ateş Ö, Alkış N. Prognostic factors and treatment outcomes in chondrosarcoma: a single-institution experience. *Anatolian Curr Med J.* 2025;7(5):657-664.

Received: 02.08.2025 • **Accepted**: 03.09.2025 • **Published**: 15.09.2025

ABSTRACT

Aims: Prognostic factors and treatment modalities in chondrosarcoma (CS) remain poorly defined. This study aimed to present our institutional experience with CS patients.

Methods: We retrospectively reviewed the medical records of 69 CS patients treated at our Cancer Institute between 2010 and 2023.

Results: Median age at diagnosis was 47 years, with a slight male predominance (52%) and a median follow-up of 93 months. Grade 1 (34%) and grade 2 (34%) tumors were most common, predominantly affecting the lower extremities (40%) and pelvis (26%). At presentation, 63 patients had localized disease and 6 had metastases; metastases later developed in 13 additional patients, most frequently in the lungs. All 63 patients with localized disease underwent surgery, and 7 received adjuvant chemotherapy (CT). Recurrence developed in 26 patients, of whom 9 were treated with systemic therapy. Among 6 patients with stage IV disease, 5 underwent palliative surgery and received CT. The most commonly used regimens included doxorubicin-based CT, pazopanib, sirolimus, and celecoxib. Among patients who developed metastasis during follow-up, 12/13 died. The median time from diagnosis to metastasis was 17.0 months (95% CI: 10.8–23.1), and median survival thereafter was 21.0 months (95% CI: 12.7–29.2). All six patients metastatic at presentation died, with a median progression-free survival of 7.0 months (95% CI: 2.1–11.8) and OS of 12.0 months (95% CI: 0.0–33.6). OS was 75%, 63%, and 57% at 5, 10, and 15 years, respectively. Female sex, low–intermediate grade, and stage I–II disease correlated with better OS in univariate analysis, but only stage at diagnosis remained significant in multivariate analysis (p=0.002).

Conclusion: Our findings highlight that early tumor stage is the only independent predictor of overall survival, underscoring the critical importance of early diagnosis and timely intervention in CS.

Keywords: Chemotherapy, surgery, radiotherapy, lung metastasis, chondrosarcoma

INTRODUCTION

Chondrosarcoma (CS) is the second most common primary skeletal tumor, with an incidence of ~3 per 100.000 person-years.[1,2] It represents a heterogeneous group with variable histopathology and clinical behavior,[3,4] ranging from locally aggressive, low-grade tumors with limited metastatic potential to high-grade malignancies with poor outcomes.[5] Conventional CS accounts for ~85% of cases, while nonconventional subtypes—clear cell, dedifferentiated (DDCS), myxoid, and mesenchymal (MCS)—comprise 10–15%.[1,6,7] Most conventional CSs (90%) are low–intermediate grade, indolent, and rarely metastasize, whereas high-grade conventional CS and rarer subtypes show high metastatic potential and poor prognosis.[8]

CS is generally refractory to chemotherapy (CT) and radiotherapy (RT), making surgery the primary treatment for low-grade, localized disease.[3] However, surgical resection may be challenging in large or anatomically complex tumors, and tumor biology can evolve, with grade discrepancies observed between biopsy, primary, and metastatic sites due to heterogeneity.[9,10] In such cases, surgery alone may be insufficient given the high risk of recurrence and metastasis,[3,5] whereas high-grade subtypes show greater responsiveness to CT and RT.

Due to the rarity of CS and its relatively long survival, prospective studies comparing prognostic factors or treatment regimens are difficult to conduct. Consequently, most evidence comes from small retrospective and single-institution series, and no consensus exists regarding prognostic factors or treatment algorithms.[11] In this study, we present our Cancer Institute's decade-long experience in managing CS, providing

Corresponding Author: Tuğba Önder, ondertugba111@gmail.com

Anatolian Curr Med J. 2025;7(5):657-664

an overview of its histopathology, classification, clinical features, treatment approaches, and prognostic factors to contribute to the limited literature.

METHODS

Ethics

The study was conducted with the permission of the Dr. Abdurrahman Yurtaslan Ankara Oncology Training and Research and Hospital Non-interventional Clinical Researches Ethics Committee (Date: 19.10.2023, Decision No: 2023-10/98). All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Study Design and Patients

The 74 patients diagnosed with CS between 2010 and 2023 in the Department of Medical Oncology of our Cancer Institute were retrospectively evaluated. Five patients were excluded from the study because of missing data. A total of 69 patients were enrolled. The American Joint Committee on Cancer (AJCC) 8 was used for staging. Patients were evaluated for clinicopathologic characteristics, treatment modalities, and prognostic factors. The primary endpoints were disease-free survival (DFS) and overall survival (OS). DFS was defined as the time from curative surgery to recurrence, and OS as the time from diagnosis to death or last follow-up. Patients without recurrence or alive at last follow-up were censored. Several clinicopathologic factors, such as sex, age, histologic subtype, stage, primary site (axial/appendicular), margin (R0/R1), and surgery, CT, and RT, were evaluated for their prognostic impact on DFS and OS.

Statistical Analysis

Continuous variables were reported as medians (range), and categorical variables as percentages. Survival was estimated using Kaplan–Meier curves, with the log-rank test applied for univariate analysis. Variables with $p<0.05$ were entered into a Cox proportional hazards model for multivariate analysis. Statistical significance was set at $p<0.05$. Analyses were performed using SPSS version 26.

RESULTS

Clinicopathologic Profile

Sixty-nine patients were included, with a median age of 47 years (range: 19–86) and a slight male predominance (53%). The most common tumor sites were the lower extremities (40%) and pelvis (26%). Grade 1 (34%) and grade 2 (34%) were the predominant histologies. At diagnosis, 63 patients (91%) had localized disease, while 6 (9%) presented with metastases. In total, 19 patients developed metastases either at presentation or during follow-up, almost all to the lungs (100%), with occasional spread to the liver (10%), thoracic spine (10%), brain (5%), or lymph nodes (5%). Patient characteristics are summarized in Table 1.

Preoperative biopsies were performed in 31 patients, but only 9 matched the final surgical pathology, reflecting a high rate of histologic upgrade or change after resection. Twelve cases initially reported as chondroid lesions, and eight as atypical chondroid tumors, were reclassified as grade 1, grade 2, or myxoid types. One biopsy diagnosed as grade 2 was upgraded to grade 3, and a lesion initially reported as a small round cell tumor was reclassified as mesenchymal CS.

Histopathological evaluation of 21 recurrent or progressive tumors showed that most (n=9) retained their initial grade, while several progressed to higher grades. Specifically, one grade 1 tumor progressed to grade 2, another to grade 3, one grade 3 became undifferentiated, and one undifferentiated tumor was reclassified as grade 3 at recurrence.

Table 1. Patient demographics and clinical characteristics (n=69)

Parameters n (%)			
Median age (range)	47 (19–86)	No	1 (1)
Gender		**Adjuvant RT**	63 (100)
Male	36 (53)	Yes	20 (31)
Female	33 (47)	No	43 (68)
Histological subtype		**Neoadjuvant/adjuvant CT**	63 (100)
Grade 1–2	48 (70)	IMA	3 (4)
Grade 3	6 (9)	NCI	3 (4)
Myxoid	5 (7)	Unknown type	2 (2)
Dedifferentiated	4 (6)	No	55 (87)
Mesenchymal	3 (4)	**Relapse of localized disease**	63 (100)
Clear cell	2 (3)	Yes	26 (41)
Unknown	1 (1)	No	37 (59)
Stage at diagnosis		**Surgery for relapse**	26 (100)
Stage 1–2	47 (68)	Yes	22 (84)
Stage 3	15 (21)	No	4 (15)
Stage 4	6 (8)	**Post-relapse RT**	26 (100)
Unknown	1 (1)	Yes	10 (38)
Tumor site		No	16 (61)
Lower extremity	28 (41)	**Post-relapse CT**	26 (100)
Upper extremity	12 (17)	Yes	9 (34)
Pelvis/ trunk/sternum	26 (38)	No	17 (65)
Mandible/maxilla	3 (4)	**Metastatic status at diagnosis**	
Preoperative biopsy		Localized	63 (91)
Yes	31 (44)	Metastatic	6 (9)
No	38 (55)	**Treatment of metastatic disease**	
Surgery		Palliative surgery	5 (83)
Curative	63 (91)	Palliative RT	4 (66)
Palliative	5 (7)	Palliative CT	5 (83)

CT: Chemotherapy, RT: Radiotherapy, IMA: Ifosfamide plus adriamycin, NCI: Vincristine, doxorubicin, cyclophosphamide, ifosfamide, etoposide

Treatment Modalities for Patients with Initially Localized Disease

The median follow-up was 93 months (range: 77–108). All 63 patients with localized disease underwent curative surgery; margin data were available in 43 patients, with 33 R0 and 10 R1 resections. Adjuvant RT was given to 20 patients. One patient received neoadjuvant CT as combination therapy of

vincristine, doxorubicin, cyclophosphamide, ifosfamide, and etoposide (NCI), and 7 received adjuvant CT, most commonly ifosfamide plus adriamycin (IMA, n=3) and NCI (n=2). CT regimens by histopathology are summarized in **Table 2**.

Recurrence and Metastasis

Recurrence occurred in 26 of 63 patients (41%): 15 had local recurrence, 8 both local and lung metastases, and 3 lung metastases alone; 2 additional patients developed lung metastases after local recurrence. In total, 13 patients developed metastases during follow-up [grade 2 (n=6), grade 3 (n=4), grade 1 (n=1), mesenchymal (n=1), dedifferentiated (n=1)], with a median time to metastasis of 17 months (95% CI: 10.8–23.1). Most recurrences (61%) occurred within 1–3 years, though 6 patients relapsed at 10–15 years. Of recurrent cases, 22 (84%) underwent surgery, 10 (38%) received RT, and 9 (34%) received CT. First-line CT most often included IMA (n=3) or cyclophosphamide adriamycin (CA) (n=6); 4 patients received second-line therapies (pazopanib, gemcitabine/docetaxel, cyclophosphamide/methotrexate/celecoxib, or ifosfamide), and 2 received third-line regimens (sirolimus/cyclophosphamide or gemcitabine/docetaxel). The median number of CT lines administered was 1 (range: 1–3).

Treatment Modalities for Patients with Initially Metastatic Disease

All six patients presenting with metastases [myxoid (n=3), dedifferentiated (n=1), mesenchymal (n=1), grade 2 (n=1)] had lung involvement. Five underwent palliative surgery and four received palliative RT. IMA/CA (n=2), IMA (n=1), CA (n=1), and NCI (n=1) were the most common CT regimens used as first-line therapy, and IMET (ifosfamide, mesna, etoposide) (n=2), gemcitabine/docetaxel (n=1), and methotrexate (n=1) were the most common regimens used as second-line therapy, while gemcitabine/docetaxel (n=1) was the regimen used as third-line therapy. Patients received a median of 2 (range: 0–3) CT lines. Single-agent therapy (n=3) yielded two progressions and one stable disease, whereas combination regimens (n=22) achieved nine stable responses and five progressions. Detailed CT responses for recurrent and metastatic patients are shown in **Table 2**.

Prognostic Factor Analysis for Disease-Free Survival and Progression-Free Survival

In patients with localized disease, DFS was 54%, 54%, and 36% at 5, 10, and 15 years, respectively. Univariate analysis showed no significant impact of sex (p=0.093), age (p=0.604), histologic subtype (p=0.326), margin status (p=0.396), adjuvant RT (p=0.118), or adjuvant CT (p=0.358) on recurrence risk. In contrast, stage at diagnosis (p=0.002) and tumor location (p=0.014) were significant prognostic factors for DFS. Detailed results are provided in **Table 3**.

Time to recurrence was significantly longer in patients with stage I–II tumors and those with appendicular tumors.

Table 2. Chemotherapy regimens by tumor histopathology and best responses to CT (n=69)

Total n (%)	**Grade 1 24 (34)**	**Grade 2 24 (34)**	**Grade 3 6 (8)**	**Myxoid 5 (7)**	**Dedifferentiated 4 (5)**	**Mesenchymal 3 (4)**	**Clear cell 2 (2)**	**Unknown 1 (1)**
Neoadjuvant						NCI 1 (1)		
Adjuvant		IMA 1 (1)	IMA 1 (1)		IMA 1 (1)	NCI 2 (2)		
Localized at diagnosis								
Recurrence 1. line	CA 1 (1) (not assessable)	IMA 2 (2) (1 progressive, 1 stable) CA 3 (3) (1progressive, 2 not assessable)	CA 2 (2) (1 stable, 1 not assessable)					IMA 1 (1) (stable)
2. line		Pazopanib 1 (1) (progressive) Gem/doc 1 (1) (progressive)	Ifosfamide 1 (1) (stable)					Cyc/mtx/ Celecoxib 1 (1) (stable)
3. line		Sirolimus/cyc 1 (1) (not assessable)	Gem/doc 1 (1) (not assessable)					
Metastatic at diagnosis								
First-line				IMA/CA 1 (1) (stable) IMA 1 (1) (stable) CA 1 (1) (progressive)	IMA/CA1 (1) (progressive)	NCI 1 (1) (stable)		
Second-line				IMET 2 (2) (2 not assessable) Gem/doc 1 (1) (stable)	Mtx 1 (1) (progressive)			
Third-line				Gem/doc 1 (1) (stable)				

CT: Chemotherapy, IMA: Ifosfamide plus adriamycin, NCI: Vincristine, doxorubicin, cyclophosphamide, ifosfamide, and etoposide, CA: cyclophosphamide, IMET: Ifosfamide, mesna, and etoposide, Gem/doc: Gemcitabine/docetaxel; MTX: Methotrexate, Cyc: Cyclophosphamide

Table 3. Prognostic factors for disease-free survival (n=63)

Factors	Univariate analysis			Multivariate analysis		
	5-yr PFS (%)	10-yr PFS (%)	p value	HR	95% CI	p value
Entire group	54	54				
Sex			0.093			
Female	66	66				
Male	40	40				
Diagnosis age			0.604			
<47 years	51	51				
>47 years	58	58				
Subtype			0.326			
Clear cell, grade1-2	56	56				
Grade3, dedifferentiated, mesenchymal, myxoid	54	54				
Stage			0.002			
Stage1	54	54		-		
Stage2	78	78		0.621	0.219-1.764	0.371
Stage3	23	23		2.658	1.070-6.605	0.035
Location			0.014			
Appendicular skeleton	66	66		-		
Axial skeleton	37	37		2.081	0.936-4.625	0.072
Surgical margin			0.396			
Negative	64	64				
Positive/closed	60	0				
Adjuvant RT			0.118			
Yes	69	69				
No	47	47				
Adjuvant CT			0.358			
Yes	43	43				
No	56	56				

RT: Radiotherapy, CT: Chemotherapy, HR: Hazard ratio, CI: Confidence interval, p values: <p–0.05.

In multivariate analysis, only stage at diagnosis remained significantly associated with DFS (p=0.013).

All patients with metastases at presentation progressed and died during follow-up, with a median time to progression of 7 months (95% CI: 2.1–11.8).

Prognostic Factor Analysis for Overall Survival

In the entire cohort, OS was 75%, 63%, and 57% at 5, 10, and 15 years, respectively. Univariate analysis identified stage at diagnosis (p<0.001), tumor grade (p=0.001), and sex (p=0.016) as significant prognostic factors for OS. Detailed results are provided in **Table 4**, and survival curves are shown in **Figure**.

Univariate analysis showed significantly longer survival in females, patients with low–moderate grade disease, stage I–II tumors, and those without metastases. In multivariate analysis, only stage at diagnosis remained an independent predictor of survival (p=0.002).

In patients with initially localized disease treated curatively, OS was 80%, 68%, and 68% at 5, 10, and 15 years, respectively. Female patients tended to have better OS than males (86%, 79%, 79% vs. 73%, 53%, 53% at 5, 10, 15 years; p=0.059), and appendicular tumors showed a survival advantage over axial tumors (84%, 80%, 80% vs. 73%, 49%, 49%; p=0.068).

Univariate analysis revealed that stage and grade significantly affected OS: stage I–II tumors had superior outcomes compared to stage III (84%, 84%, 84% vs. 90%, 90%, 90% vs. 58%, 14%, 0% at 5, 10, and 15 years; p<0.001), and low–intermediate grade tumors outperformed high-grade tumors (83%, 72%, 72% vs. 66%, 44%, 44% at 5, 10, 15 years; p=0.026). In multivariate analysis, only stage at diagnosis remained independently significant, with stage I tumors showing a markedly reduced risk of death compared to stage IV (HR=0.099, 95% CI: 0.1–0.5; p=0.011).

In the initially metastatic group, median OS was 12 months (95% CI: 0.1–33.6). The development of metastasis during follow-up was a significant predictor of OS (p=0.001; HR=0.064, 95% CI: 0.1–0.3).

DISCUSSION

The aim of this study was to evaluate the clinicopathological features, treatment modalities, prognoses, and outcomes of patients with CS. We presented a comprehensive overview of all CS cases with varying histological subtypes treated in our

Table 4. Prognostic factors for overall survival (n=69)

Factors	Univariate analysis			Multivariate analysis		
	5-yr OS (%)	10-yr OS (%)	p value	HR	95% CI	p value
Entire group	75	63				
Sex			0.016			
Female	83	77		-		
Male	64	48		1.541	0.554-4.286	0.407
Diagnosis age			0.936			
<47 years	75	58				
>47 years	75	57				
Subtype			0.001			
Clear cell, grade1-2	92	85		-		
Grade3, dedifferentiated, mesenchymal, myxoid	52	37		1.325	0.426-4.126	0.627
Stage			<0.001			
Stage1	90	85		0.099	0.017-0.589	0.011
Stage2	89	89		0.068	0.011-0.420	0.004
Stage3	58	14		0.575	0.166-1.991	0.383
Stage4	17	0		-		
Location			0.125			
Appendicular skeleton	78	74				
Axial skeleton	68	45				
Surgical margin			0.101			
Negative	93	70				
Positive/closed	53	0				

HR: Hazard ratio, CI: Confidence interval, p values: <p–0.05.

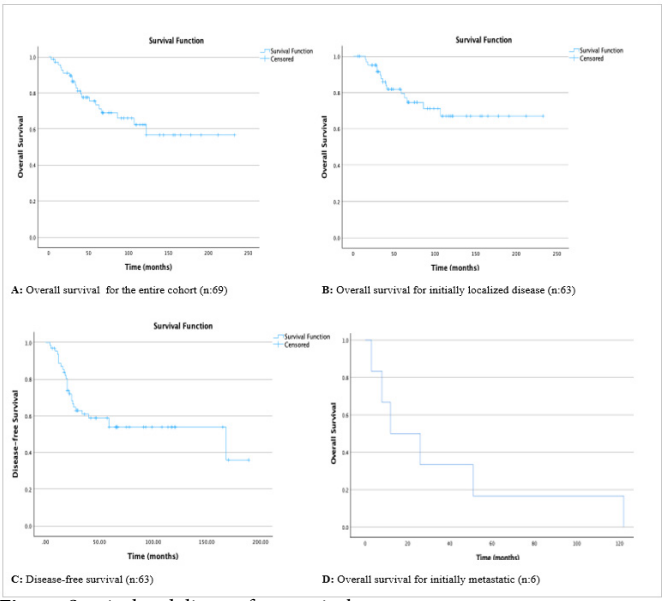


Figure. Survival and disease-free survival curves

clinic. By doing so, we aimed to contribute to the limited body of literature on this rare malignancy and provide additional insights into its management and outcomes.

In the present study, the OS rates at 5, 10, and 15 years were 75%, 63%, and 57%, respectively, consistent with previous reports.[4,12-14] The median age at diagnosis was 47 years, with a slight male predominance (52%), whereas earlier studies reported a median age of ~50 years and nearly equal sex distribution.[4,13,15] Male sex was associated with poorer prognosis in univariate analysis, in line with prior studies,[13,15] although some reports found no significant sex-related survival difference.[11,16]

Histological grade is a major prognostic factor in CS.[8] In our cohort, nearly all metastatic cases were intermediate- or high-grade, and tumor grade significantly influenced OS, consistent with the literature.[11,13] Low-grade CS rarely metastasizes,[4,5,13] but once present, outcomes are poor, often reflecting progression to a more aggressive phenotype.[16-19] In the current study, notably, one patient with an initial low-grade tumor developed metastasis that had transformed to grade 3.

The development of local recurrence or metastasis was associated with significantly worse survival. Most recurrences in our cohort occurred within 1–3 years, but six patients relapsed 10–15 years after diagnosis, consistent with reports of recurrences up to 20 years.[20-22] Although conventional and dedifferentiated histologies were observed in these late cases, the limited sample size precluded firm conclusions about tumor biology and late relapse. These findings emphasize the importance of close surveillance in the first five years and continued long-term follow-up due to the potential for late relapse.

An atypical cartilaginous tumor (ACT), formerly termed grade I CS, refers to tumors of the appendicular skeleton and reflects distinct biological behavior by site.[8] ACTs grow

slowly, behave locally aggressively, and rarely metastasize.[23] In our study, local recurrence occurred in 26% of appendicular and 33% of axial grade I tumors. Overall, axial CSs tend to have poorer outcomes and are treated more aggressively,[3,5,24] although some reports suggest location does not significantly affect survival.[3,4,13] In our cohort (22 axial, 47 appendicular cases), location was a significant prognostic factor for DFS but not OS in univariate analysis. This may relate to the tumor microenvironment, treatment differences by site, or delayed diagnosis due to anatomical constraints.[15]

In our study, stage at diagnosis significantly influenced both DFS and OS in multivariate analysis, with stage II patients showing slightly better outcomes than stage I. Andreou et al.[16] reported no correlation between AJCC stage and outcomes in localized axial or pelvic CS. Notably, histological grading may vary among pathologists,[25] which can alter staging and affect surgical decisions. Larger studies are needed to compare and validate staging systems.

In our cohort, all 19 patients with metastatic disease at diagnosis or follow-up had lung involvement. Similar to other sarcomas, the lung was the predominant metastatic site, with far less frequent spread to the liver, vertebrae, lymph nodes, or brain. Brain metastases are extremely rare, with only 12 cases reported in the literature;[26] in our series, one patient with grade II CS developed lung, brain, and thoracic vertebral metastases. Primary spinal CS is also uncommon[27] and may present with spinal cord compression. In our study, one patient with stage IV mesenchymal-type CS developed spinal paraplegia and received palliative RT for spinal metastasis.

Biopsy is useful for diagnosis and surgical planning but may not accurately determine histological grade due to heterogeneity, sampling error, or interobserver variability.[25,28,29] In our study, preoperative biopsy results matched surgical pathology in only 9 of 31 patients, underscoring the need to integrate clinical, radiological, and histological findings.[30] Recurrent CSs may present with higher grades than the primary tumor;[5,31,32] in our series, 16 of 21 recurrent cases retained the same grade. Some CSs also show unpredictable behavior despite appropriate classification.[13,33,34] Thus, molecular markers and advanced imaging are needed to better predict prognosis, guide therapy, and improve preoperative diagnosis.[30,35,36]

Adequate surgical excision is critical in CS management. In our cohort, patients with negative margins showed a slight, though non-significant, survival advantage, likely limited by sample size. Previous studies, however, have demonstrated significant differences in OS and DFS based on margin status.[4,16]

Low-grade CSs are generally resistant to RT and CT due to their slow growth and low mitotic activity.[5,6,37,38] Nevertheless, RT may be considered for unresectable or borderline cases and for palliation,[37,39,40] while CT shows activity in mesenchymal and dedifferentiated subtypes.[6,7,41] Given the rarity of CS, evidence is limited to retrospective series.[6,37,42] In our study, IMA, NCI, and CA were the most frequently used regimens. CT was mainly administered to mesenchymal, dedifferentiated, or grade III tumors, where higher proliferative potential and aggressive course suggest greater benefit.

In our study, CT was administered across all histologies except clear cell. Only three patients received single-agent therapy, while the rest received combinations. Among recurrent and metastatic cases, combination regimens showed better disease control (5 progressions/9 stable disease) compared to single agents (2 progressions/1 stable disease), though small numbers precluded subtype-specific conclusions. Similarly, Italiano et al.[42] reported higher response rates with combination therapy than with single agents (20% vs. 11%, p=0.09).

Interest in targeted and immunotherapies for advanced or unresectable CS has been increasing, given the limited efficacy of CT and RT. Molecular studies highlight potential targets such as IDH1/2 mutations, hedgehog signaling, and angiogenesis pathways.[2,30] Early-phase trials of IDH inhibitors (e.g., ivosidenib) have shown promise in IDH-mutant CS,[43] while pazopanib has achieved disease stabilization in advanced cases.[1] In contrast, checkpoint inhibitors have yielded limited benefit, likely due to the immunologically "cold" tumor microenvironment,[44] though ongoing studies are testing novel combinations to improve immune responsiveness.[45,46]

Limitations

This study has several limitations. Missing data on factors such as comorbidities, surgical margin width, and RT dose may have influenced outcomes, as these are prognostically relevant in CS.[13,23,33] Competing risks, such as death from other causes, may also affect survival analyses.[13,47,48] Moreover, the long study period may have introduced variability in treatment approaches. The retrospective design, rarity of CS, small sample size, heterogeneity of treatment periods, and reliance on single-center data without external validation limit the generalizability of our findings. Future multicenter studies with larger cohorts are needed to validate and generalize these findings.

CONCLUSION

As a result, this study adds to the limited data on the prognosis of local and metastatic CS. While no novel findings were identified, our results confirm prior knowledge by showing that stage and tumor location predict DFS, whereas stage, sex, and histological subtype influence OS. Multivariate analysis indicated that early tumor stage was the only independent prognostic factor for both DFS and OS. Given the heterogeneity of CS, a multidisciplinary approach with standardized criteria is essential to optimize diagnosis and treatment. The regional longitudinal dataset in our study offers important insights for the clinical management of CS and serves as a valuable reference for future comparative studies.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was conducted with the permission of the Dr. Abdurrahman Yurtaslan Ankara Oncology Training and Research and Hospital Non-interventional Clinical Researches Ethics Committee (Date: 19.10.2023, Decision No: 2023-10/98).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Chow W, Frankel P, Ruel C, et al. Results of a prospective phase 2 study of pazopanib in patients with surgically unresectable or metastatic chondrosarcoma. *Cancer*. 2020;126(1):105-111. doi:10.1002/cncr.32515
2. van Oosterwijk JG, Anninga JK, Gelderblom H, Cleton-Jansen AM, Bovee JV. Update on targets and novel treatment options for high-grade osteosarcoma and chondrosarcoma. *Hematol Oncol Clin North Am*. 2013;27(5):1021-1048. doi:10.1016/j.hoc.2013.07.012
3. Angelini A, Guerra G, Mavrogenis AF, Pala E, Picci P, Ruggieri P. Clinical outcome of central conventional chondrosarcoma. *J Surg Oncol*. 2012;106(8):929-937. doi:10.1002/jso.23173
4. Fiorenza F, Abudu A, Grimer RJ, et al. Risk factors for survival and local control in chondrosarcoma of bone. *J Bone Joint Surg Br*. 2002;84(1):93-99. doi:10.1302/0301-620x.84b1.11942
5. Evans HL, Ayala AG, Romsdahl MM. Prognostic factors in chondrosarcoma of bone: a clinicopathologic analysis with emphasis on histologic grading. *Cancer*. 1977;40(2):818-831. doi:10.1002/1097-0142(197708)40:2<818::aid-cncr2820400234>3.0.co;2-b
6. Gelderblom H, Hogendoorn PC, Dijkstra SD, et al. The clinical approach towards chondrosarcoma. *Oncologist*. 2008;13(3):320-329. doi:10.1634/theoncologist.2007-0237
7. Kim JH, Lee SK. Classification of chondrosarcoma: from characteristic to challenging imaging findings. *Cancers (Basel)*. 2023;15(6):1703. doi:10.3390/cancers15061703
8. Choi JH, Ro JY. The 2020 WHO classification of tumors of bone: an updated review. *Adv Anat Pathol*. 2021;28(3):119-138. doi:10.1097/PAP.0000000000000293
9. Fromm J, Klein A, Baur-Melnyk A, et al. Survival and prognostic factors in conventional central chondrosarcoma. *BMC Cancer*. 2018;18(1):849. doi:10.1186/s12885-018-4741-7
10. van Praag Veroniek VM, Rueten-Budde AJ, Ho V, et al. Incidence, outcomes and prognostic factors during 25 years of treatment of chondrosarcomas. *Surg Oncol*. 2018;27(3):402-408. doi:10.1016/j.suronc.2018.05.009
11. Amer KM, Munn M, Congiusta D, Abraham JA, Basu Mallick A. Survival and prognosis of chondrosarcoma subtypes: SEER database analysis. *J Orthop Res*. 2020;38(2):311-319. doi:10.1002/jor.24463
12. Dorfman HD, Czerniak B. Bone cancers. *Cancer*. 1995;75(1 Suppl): 203-210. doi:10.1002/1097-0142(19950101)75:1+<203::aid-cncr2820751308>3.0.co;2-v
13. Giuffrida AY, Burgueno JE, Koniaris LG, Gutierrez JC, Duncan R, Scully SP. Chondrosarcoma in the United States (1973 to 2003): an analysis of 2890 cases from the SEER database. *J Bone Joint Surg Am*. 2009;91(5):1063-1072. doi:10.2106/JBJS.H.00416
14. Damron TA, Ward WG, Stewart A. Osteosarcoma, chondrosarcoma, and Ewing's sarcoma: National Cancer Data Base Report. *Clin Orthop Relat Res*. 2007;459:40-47. doi:10.1097/BLO.0b013e318059b8c9
15. Bindiganavile S, Han I, Yun JY, Kim HS. Long-term outcome of chondrosarcoma: a single institutional experience. *Cancer Res Treat*. 2015;47(4):897-903. doi:10.4143/crt.2014.135
16. Andreou D, Ruppin S, Fehlberg S, Pink D, Werner M, Tunn PU. Survival and prognostic factors in chondrosarcoma: results in 115 patients with long-term follow-up. *Acta Orthop*. 2011;82(6):749-755. doi:10.3109/17453674.2011.636668
17. Leerapun T, Hugate RR, Inwards CY, Scully SP, Sim FH. Surgical management of conventional grade I chondrosarcoma of long bones. *Clin Orthop Relat Res*. 2007;463:166-172. doi:10.1097/BLO.0b013e318146830f
18. Schwab JH, Wenger D, Unni K, Sim FH. Does local recurrence impact survival in low-grade chondrosarcoma of the long bones? *Clin Orthop Relat Res*. 2007;462:175-180. doi:10.1097/BLO.0b013e3180caac2c
19. Streitburger A, Ahrens H, Balke M, et al. Grade I chondrosarcoma of bone: the Munster experience. *J Cancer Res Clin Oncol*. 2009;135(4):543-550. doi:10.1007/s00432-008-0486-z
20. Itala A, Leerapun T, Inwards C, Collins M, Scully SP. An institutional review of clear cell chondrosarcoma. *Clin Orthop Relat Res*. Nov 2005;440:209-212. doi:10.1097/01.blo.0000174686.70003.e0
21. Donati D, Yin JQ, Colangeli M, et al. Clear cell chondrosarcoma of bone: long time follow-up of 18 cases. *Arch Orthop Trauma Surg*. 2008;128(2):137-142. doi:10.1007/s00402-007-0353-4
22. Dickey ID, Rose PS, Fuchs B, et al. Dedifferentiated chondrosarcoma: the role of chemotherapy with updated outcomes. *J Bone Joint Surg Am*. 2004;86(11):2412-2418.
23. Wang M, Song Y, Liu S, Sun W. Effect of surgery and radiotherapy on overall survival in patients with chondrosarcoma: a SEER-based study. *J Orthop Surg (Hong Kong)*. 2022;30(1):10225536221086319. doi:10.1177/10225536221086319
24. Donati D, El Ghoneimy A, Bertoni F, Di Bella C, Mercuri M. Surgical treatment and outcome of conventional pelvic chondrosarcoma. *J Bone Joint Surg Br*. 2005;87(11):1527-1530. doi:10.1302/0301-620X.87B11.16621
25. Eefting D, Schrage YM, Geirnaerdt MJ, et al. Assessment of interobserver variability and histologic parameters to improve reliability in classification and grading of central cartilaginous tumors. *Am J Surg Pathol*. 2009;33(1):50-57. doi:10.1097/PAS.0b013e31817eec2b
26. Yokota A, Ogiwara T, Hanaoka Y, Higashiyama F, Oya F, Horiuchi T. Brain metastasis of chondrosarcoma: a case report and literature review. *Radiol Case Rep*. 2023;18(5):1939-1944. doi:10.1016/j.radcr.2023.02.040
27. Qubain L, Hirsch BP, Reddivalla N, Baaj AA, Leddy LR, Ravinsky RA. Metastatic mesenchymal chondrosarcoma of the spine managed with nonsurgical treatment: a case report and review of the literature. *Clin Case Rep*. 2022;10(12):e6712. doi:10.1002/ccr3.6712
28. Normand AN, Cannon CP, Lewis VO, Lin PP, Yasko AW. Curettage of biopsy-diagnosed grade 1 periacetabular chondrosarcoma. *Clin Orthop Relat Res*. 2007;459:146-149. doi:10.1097/BLO.0b013e3180619554
29. Skeletal Lesions Interobserver Correlation among Expert Diagnosticians Study G. Reliability of histopathologic and radiologic grading of cartilaginous neoplasms in long bones. *J Bone Joint Surg Am*. 2007;89(10):2113-2123. doi:10.2106/JBJS.F.01530
30. Bovee JV, Cleton-Jansen AM, Taminiau AH, Hogendoorn PC. Emerging pathways in the development of chondrosarcoma of bone and implications for targeted treatment. *Lancet Oncol*. 2005;6(8):599-607. doi:10.1016/S1470-2045(05)70282-5
31. Grimer RJ, Gosheger G, Taminiau A, et al. Dedifferentiated chondrosarcoma: prognostic factors and outcome from a European group. *Eur J Cancer*. 2007;43(14):2060-2065. doi:10.1016/j.ejca.2007.06.016
32. Björnsson J, McLeod RA, Unni KK, Ilstrup DM, Pritchard DJ. Primary chondrosarcoma of long bones and limb girdles. *Cancer*. 1998;83(10):2105-2119.
33. Eriksson AI, Schiller A, Mankin HJ. The management of chondrosarcoma of bone. *Clin Orthop Relat Res*. 1980;153:44-66.
34. Mankin HJ, Cantley KP, Lippiello L, Schiller AL, Campbell CJ. The biology of human chondrosarcoma. I. description of the cases, grading, and biochemical analyses. *J Bone Joint Surg Am*. 1980;62(2):160-176.
35. Sharif B, Rajakulasingam R, Sharifi S, O'Donnell P, Saifuddin A. MRI features of low-grade and high-grade chondrosarcoma in enchondromatosis. *Skeletal Radiol*. 2021;50(8):1637-1646. doi:10.1007/s00256-021-03718-7
36. Li X, Lan M, Wang X, et al. Development and validation of a MRI-based combined radiomics nomogram for differentiation in chondrosarcoma. *Front Oncol*. 2023;13:1090229. doi:10.3389/fonc.2023.1090229

37. van Maldegem AM, Gelderblom H, Palmerini E, et al. Outcome of advanced, unresectable conventional central chondrosarcoma. *Cancer*. 2014;120(20):3159-3164. doi:10.1002/cncr.28845
38. Le A, Ball D, Pitman A, Fox R, King K. Chondrosarcoma of bone complicating Ollier's disease: report of a favourable response to radiotherapy. *Australas Radiol*. 2003;47(3):322-324. doi:10.1046/j.1440-1673.2003.01187.x
39. Catanzano AA, Kerr DL, Lazarides AL, et al. Revisiting the role of radiation therapy in chondrosarcoma: a national cancer database study. *Sarcoma*. 2019;2019:4878512. doi:10.1155/2019/4878512
40. McNaney D, Lindberg RD, Ayala AG, Barkley HT Jr, Hussey DH. Fifteen year radiotherapy experience with chondrosarcoma of bone. *Int J Radiat Oncol Biol Phys*. 1982;8(2):187-190. doi:10.1016/0360-3016(82)90512-0
41. Frezza AM, Cesari M, Baumhoer D, et al. Mesenchymal chondrosarcoma: prognostic factors and outcome in 113 patients. A European Musculoskeletal Oncology Society study. *Eur J Cancer*. 2015; 51(3):374-381. doi:10.1016/j.ejca.2014.11.007
42. Italiano A, Mir O, Cioffi A, et al. Advanced chondrosarcomas: role of chemotherapy and survival. *Ann Oncol*. 2013;24(11):2916-2922. doi:10.1093/annonc/mdt374
43. Tap WD, Cote GM, Burris H, et al. Phase I study of the mutant IDH1 inhibitor ivosidenib: long-term safety and clinical activity in patients with conventional chondrosarcoma. *Clin Cancer Res*. 2025;31(11):2108-2114. doi:10.1158/1078-0432.CCR-24-4128
44. Wagner MJ, Pimenta EM, Sweeney NW, et al. Genomic characterization of chondrosarcoma reveals potential therapeutic targets. *JCO Precis Oncol*. 2025;9:e2400592. doi:10.1200/PO-24-00592
45. Chua-Alcala VS, Kim K, Assudani N, et al. Initial results of a phase III investigation of safety/efficacy of nivolumab and ABI-009 (nab-sirolimus) in advanced undifferentiated pleomorphic sarcoma (UPS), liposarcoma (LPS), chondrosarcoma (CS), osteosarcoma (OS), and Ewing sarcoma. *Am Society Clin Oncol*. 2019;37(Suppl 8):21. doi:10.1200/JCO.2019.37.8_suppl.21
46. Cohen-Nowak AJ, Dressler DB, Rock A, et al. Role of immunotherapy in chondrosarcoma: a case report and review of the literature. *Ther Adv Med Oncol*. 2023;15:17588359231199877. doi:10.1177/17588359231199877
47. Pintilie M. Analysing and interpreting competing risk data. *Stat Med*. 2007;26(6):1360-1367. doi:10.1002/sim.2655
48. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med*. 1999;18(6):695-706. doi:10.1002/(sici)1097-0258(19990330)18:6<695::aid-sim60>3.0.co;2-o

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1721428

Anatolian Curr Med J. 2025;7(5):665-669

Can seasonal differences affect the frequency and mortality of pulmonary embolism?

Şule Esen Karamişe[1], Şule Çilekar[2], Aydın Balcı[2], Gökhan Karamişe[3], Sibel Günay[4], Ersin Günay[5]

[1]Department of Chest Diseases, Afyonkarahisar State Hospital, Afyonkarahisar, Turkiye
[2]Department of Chest Diseases, Faculty of Medicine, Afyonkarahisar Health Sciences University, Afyonkarahisar, Turkiye
[3]Department of Emergency Medicine, Afyonkarahisar State Hospital, Afyonkarahisar, Turkiye
[4]Department of Chest Diseases, Ankara Bilkent City Hospital, Ankara, Turkiye
[5]Department of Chest Diseases, Ankara Etlik City Hospital, Ankara, Turkiye

Cite this article as: Karamişe ŞE, Çilekar Ş, Balcı A, Karamişe G, Günay S, Günay E. Can seasonal differences affect the frequency and mortality of pulmonary embolism? *Anatolian Curr Med J.* 2025;7(5):665-669.

Received: 18.06.2025 ◆ **Accepted**: 04.09.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: Does a seasonal variation in the incidence of pulmonary embolism (PE) exist? The aim of our study is to assess seasonal variation in the incidence of PE and to determine whether there are any differences in mortality.

Methods: Medical documentation of 402 patients diagnosed as acute PE in the five-year period were retrospectively analyzed. The study included patients aged 18 years and older with a diagnosis of PE confirmed by spiral thorax computed tomography.

Results: The highest numbers of case were seen in spring (n=115, 28.6%). In terms of months, the greatest number of cases occurred in November (n=48, 11.9%) and April (n=47, 11.7%). Case distribution according to seasons and months were statistically significant (p<0.001). No seasonal differences in mortality rates were observed (p=0.663). Only the distribution of white blood cell (WBC) level by the seasons was statistically significant among the recorded numerical data (p=0.045). The differences in mortality rates by numerical data (age, WBC, neutrophil, lymphocyte, platelet, D-dimer, CRP, PAB values, sPESI and Wells scores) were statistically significant.

Conclusion: Our observation that seasonal differences affective on frequency of PE has been confirmed.

Keywords: Pulmonary embolism, seasons, mortality

INTRODUCTION

Pulmonary embolism (PE) is an obstruction of the pulmonary artery and/or its branches by fragments detached from thrombi formed in the venous system, hindering blood flow. Most PE cases originate from thrombi in the deep veins of the lower extremities.[1] Major trauma, surgery, lower-limb fractures and joint replacements, and spinal cord injury are strong provoking factors for VTE. Cancer is a well-recognized predisposing factor for VTE. The impact of meteorological conditions on cardiovascular morbidity and mortality has long been investigated. Does a seasonal variation in the incidence of PE exist? This study aims to investigate the effects of seasonal changes on the frequency and mortality of PE, motivated by the observed periodic differences in the incidence of PE cases presenting to our clinic or consulted from other departments.

METHODS

Our study was approved by the Afyonkarahisar Health Sciences University Non-interventional Clinical Researches Ethics Committee (Date: 03.05.2019, Decision No: 2019/7-173). All information and data were archieved in an encoded form to ensure confidentiality and compliance with ethical standards. All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

In our study, data from 402 patients diagnosed with acute PE by the Pulmonology Clinic Afyonkarahisar Health Sciences University Faculty of Medicine between January 2014 and January 2019 (patients presenting to our clinic or consulted from other departments) were retrospectively analyzed. Patients aged 18 and over, with PE diagnosis confirmed by spiral thorax computed tomography (CT), were included in the study. The study excluded patients younger than 18 years of age as well as those whose diagnosis was not confirmed by CT. Patients' age and gender, months of presentation, radiological classification of PE as massive or non-massive, risk factors for PE development, main laboratory parameters at presentation (white blood cell (WBC), neutrophil,

Corresponding Author: Şule Esen Karamişe, suleesen@gmail.com

lymphocyte, hemoglobin, platelet, D-dimer, CRP), pulmonary artery pressure (PAP) values, Wells scores, probabilities of PE development according to Wells scores, sPESI scores, treatments, and outcomes were recorded. The distribution of continuous variables was determined using the Kolmogorov Smirnov test. Continuous variables with normal distribution were expressed as mean±standard deviation (SD), while those without normal distribution were expressed as median (minimum-maximum). Categorical variables were expressed as number of patients (n) and percentage (%). Comparisons between two groups were made using the Student t-test for continuous variables with normal distribution, and the Mann-Whitney U test for variables without normal distribution. Post-hoc analyses (Benferroni, Tukey) were applied to variables found to be statistically significant in multi-group analyses. The Chi-square test was used to compare proportions between groups. A p-value of <0.05 was considered statistically significant.

RESULTS

In our study, a total of 402 patients diagnosed with PE were included. Of these, 208 were female (51.7%) and 194 were male (48%), with a median age of 67 years, ranging from 19 to 93 years old. When examining the seasonal frequency of PE development, the highest number of cases was observed in the spring (n=115, 28.6%), followed by autumn (n=99, 25%). The lowest number of cases was equally recorded in winter and summer (n=94, 23.4%) (**Figure 1**). The differences in the frequency of PE development across seasons were found to be statistically significant (p<0.001).

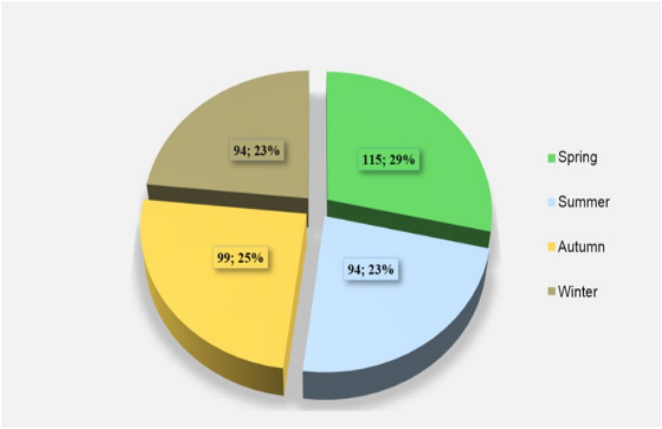


Figure 1. Frequency of pulmonary embolism development according to seasons

Regarding the monthly distribution of PE cases, the highest number were observed in November (n=48, 11.9%) and April (n=47, 11.7%), with the lowest in February and October (n=25, 6.2%) (**Figure 2, Table 1**). The variations in the frequency of PE development across months were statistically significant (p<0.001).

Seasonal distribution of risk factors highlighted in **Table 2**.

When assessing prognosis by season, differences in mortality rates were observed but were not statistically significant (p=0.663). When the prognosis was evaluated according to months, it was found that mortality rate distributions

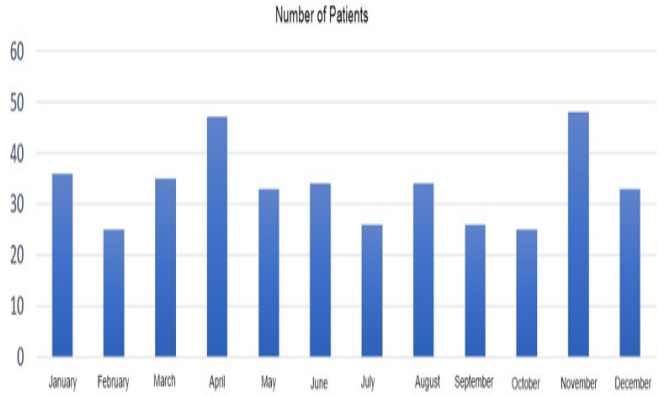


Figure 2. Frequency of pulmonary embolism development according to months

Table 1. Number of patients by month

Seasons	n	%
January	36	9
February	25	6.2
March	35	8.7
April	47	11.7
May	33	8.2
June	34	8.5
July	26	6.5
Agust	34	8.5
September	26	6.5
October	25	6.2
November	48	11.9
December	33	8.2

Table 2. Seasonal distribution of risk factors

Risk factor	Spring n (%)	Summer n (%)	Autumn n (%)	Winter n (%)	p-value
Malignancy	a	25 (26.6)	21 (21.2)	23 (24.5)	0.56
Surgical	7 (6.1)	10 (10.6)	7 (7.1)	8 (8.5)	0.656
Lower extremity	10 (8.7)	12 (12.8)	9 (9.1)	12 (12.8)	0.661
Heart disease	19 (16.5)	13 (13.8)	15 (15.2)	15 (16.0)	0.957
Immobility	17 (14.8)	12 (12.8)	20 (20.2)	13 (13.8)	0.482
Advanced age	14 (12.2)	4 (4.3)	7 (7.1)	12 (12.8)	0.115
Other	14 (12.2)	18 (19.1)	20 (20.2)	11 (11.7)	0.205

Percentages denote the seasonal distribution of each risk factor

showed differences, but these differences were not statistically significant (p=0.582). The relationships between patients' age, WBC, neutrophil, lymphocyte, platelet count, D-dimer, CRP, PAP values, PESI(s), and Wells scores with prognosis were found to be statistically significant and are summarized in **Table 3**. No significant relationship was found between hemoglobin and prognosis.

Of the recorded numerical data, only the WBC levels showed a statistically significant variation by season (p=0.045), indicating an increase in WBC levels during the summer (**Table 4**). The relationship between the seasons and the probability of embolism development according to Wells score; When the relationship between the seasons and the

Table 3. Relationship between numerical data and prognosis in PE patients

Numerical data	Alive mean (min-max)	Dead mean (min-max)	p value
Age	66.00 (19.00-93.00)	79.00 (38.00-90.00)	<0.001
WBC	10.50 (2.20-34.00)	13.00 (6.20-27.90)	0.001
Neutrophil	7.70 (0.70-32.00)	10.4 (2.90-27.00)	<0.001
Lymphocyte	1.50 (0.10-10.00)	1.00 (0.20-10.00)	0.014
Hb *	12.40±1.95	12.25±1.99	0.611
PLT	253.00 (52.00-671.00)	207.00 (58.00-831.00)	0.004
D-dimer	3.00 (0.50-42.00)	6.00 (1.40-38.60)	<0.001
CRP	5.00 (0.10-33.00)	12.70 (0.20-41.00)	<0.001
PAP	40.00 (15.00-100.00)	60.00 (24.00-90.00)	<0.001
PESI(s)	1.00 (0.00-3.00)	3.00 (0.00-4.00)	<0.001
Wells	4.50 (0.00-10.00)	6.00 (1.00-10.00)	<0.001

*Hb is normally distributed, its given as mean±standard deviation. PE: Pulmonary embolism, Min: Minimum, Max: Maximum, WBC: White blood cell, Hb: Hb: Hemoglobin, PLT: Platelet, CRP: C-reactive protein, PAP: Pulmonary artery pressure

probability of embolism development according to Wells score was examined, no statistically significant difference was found (p=0.851). Among the PE patients examined, 117 (29.1%) were assessed as having massive PE and 285 (70.9%) as non-massive. The changes in radiological classification between massive and non-massive PE by season were not statistically significant (p=0.421).

DISCUSSION

The main finding from our study is that a seasonal variation in the incidence of PE exists and that meteorological factors may contribute to this variation. Our study found the highest number of PE patients in the spring (n=115). This was followed by autumn (n=99), with the lowest equally in winter and summer (n=94). In a study by Meral et al.,[5] an increase in PE cases was observed in the spring compared to other seasons. In a thesis study at Trakya University's Pulmonology Clinic,[6] an increase in PE patients in the spring was also found to be statistically significant, but in contrast to our study, the lowest number of patients was observed in autumn. Öztuna et al.,[7] similarly found the highest number of patients in the spring. Contrary to our findings, Anar et al.[8] reported the highest number of patients in autumn and the lowest in winter. In another study in Turkiye by Ercan et al.,[9] an increase in PE development was observed in winter. In a multicenter study by Manfredini et al.[10] in Italy with 2.119 patients, the lowest number of VTE cases was in spring, and the highest in winter, contrary to our study's findings. In a study conducted by Gallerani et al.[11] in Italy, which involved a total of 19.245 patients, it was found that the frequency of PE development decreased in the spring and increased in the winter. A study by Guijarro et al.[12] in Spain included a total of 162.032 PE patients, finding an increase in PE incidence during the winter months, with the least number of cases recorded in the summer. While some studies investigating the relationship between the seasonal variation and the incidence of PE have found significant differences, others, such as the study by Stein et al.,[13] have not. Our findings add to the growing evidence supporting a seasonal variation in PE. However, controversy still exists, mainly as a result of conflicting results.

In our study, the highest number of PE patients were observed in November (n=48) and April (n=47). Öztuna et al.,[7] in their study, found the highest number of patients in May and April. In the study by Anar et al.,[8] the most PE cases were recorded in June and November. The seasonal variation in patients' radiological presentation as massive or non-massive was not found to be statistically significant. Our study found that only the seasonal variations in WBC levels were statistically significant, with an increase noted in the summer months. However, determining a single underlying cause for this observation is difficult. The highlights the need for more comprehensive and prospective studies to further clarify the relationship. While there are studies in the literature that have found significant increases in some inflammatory markers in PE patients, sufficient studies examining these increases in relation to seasonal variations are lacking. Our study found that laboratory data, scoring systems, and PAP values were significantly related to prognosis. The only laboratory parameter not significantly associated with mortality was hemoglobin distribution.

Lower platelet levels have been observed in deceased patients. An increase in platelet levels during warmer months

Table 4. Seasonal distribution of numerical data evaluated in patients with pulmonary embolism

Numerical data	Spring mean (min-max)	Summer mean (min-max)	Autumn mean (min-max)	Winter mean (min-max)	p value
Age	68.00 (21.00-90.00)	65.00 (26.00-89.00)	66.00 (23.00-93.00)	68.00 (19.00-92.00)	0.303
WBC	10.00 (2.20-24.00)¥	12.15 (2.30-34.00)¥,µ	11.30 (3.90-27.00)	10.00 (2.90-24.00)µ	0.045
Neutrophil	7.70 (1.30-23.60)	9.35 (0.70-32.00)	8.60 (2.00-23.60)	7.60 (1.60-22.00)	0.115
Lymphocyte	1.50 (0.20-4.20)	1.40 (0.10-7.60)	1.50 (0.20-10.00)	1.30 (0.20-6.10)	0.835
Hb *	12.53±1.92	12.11±2.03	12.28±1.94	12.57±1.94	0.308
PLT	233.00 (58.00-529.00)	268.00 (84.00-578.00)	253.00 (57.00-831.00)	246.50 (52.00-651.00)	0.185
D-dimer	3.00 (0.60-36.00)	3.00 (0.50-41.00)	4.00 (0.80-38.60)	3.00 (0.60-42.00)	0.141
CRP	5.00 (0.10-32.00)	6.00 (0.20-40.00)	5.80 (0.40-38.00)	7.65 (0.40-41.00)	0.398
PAP	40.00 (15.00-100.00)	40.00 (15.00-80.00)	40.00 (15.00-90.00)	45.00 (17.00-80.00)	0.307
PESI(s)	1.00 (0.00-4.00)	1.00 (0.00-4.00)	1.00 (0.00-4.00)	1.00 (0.00-4.00)	0.852
Wells	4.50 (1.50-10.00)	4.50 (1.00-9.00)	4.50 (0.00-10.00)	4.50 (1.00-9.00)	0.499

Min: Minimum, Max: Maximum, WBC: White blood cell, Hb: Hb: Hemoglobin, PLT: Platelet, CRP: C-reactive protein, PAP: Pulmonary artery pressure, *Hb is normally distributed, its given as mean±standard deviation. Post hoc analysis results of parameters with the same superscript symbol; ¥, p=0.013;µ, p=0.023

was observed, but these differences were not statistically significant. The lower platelet levels in other seasons compared to summer, and the higher number of deaths resulting from PE in these months, could be associated with a consumption-related decrease in platelets in patients with larger thrombi and massive PE.

Contrary to our findings, a study by Masotti et al.[14] found a significant increase in platelet levels during colder months, which they explained as a mild surface cooling effect increasing red blood cell and platelet counts. In this study, this was explained 'by the fact that slight surface cooling increased the number of red blood cells and platelets.' The reasons underlying the seasonal variations in the development of PE have not been fully explained. Changes in atmospheric pressure and temperatures have been suggested to be effective. Some studies have shown that cold weather is associated with blood viscosity and platelet levels. In addition, cold weather often leads to a more sedentary lifestyle which may increase thrombotic tendencies. However, during hotter days and holiday periods individuals are more likely to undertake prolonged travel, particularly by car, which may also elevated the risk. Other important risk factors may include prothrombotic changes associated with seasonal conditions, such as acute respiratory infections. The incidence of respiratory infections is known to vary throughout the year.

Seasonal variations in some VTE risk factors could also contribute to the seasonal differences in PE frequency. Accompanying pulmonary heart diseases (KPH) may also vary with the seasons. Seasonal changes in the number and types of surgical operations may also affect the frequency of PE development.

Limitations

When examining the limitations of our study, we note that some studies investigating the seasonal frequency of PE development also consider barometric pressure, temperature, and humidity values, which were not included in our study. Potential limitations should be considered when interpreting the results of the present study. The study was retrospective, and this may raise concern over the possibility of missed cases. Although our sample size was sufficient, larger population studies with higher numbers could potentially find significant seasonal differences.

CONCLUSION

Our observation that seasonal differences affect the frequency of PE has been confirmed. In our study, the highest number of cases were seen in spring. In light of this information, PE should primarily be considered in patients presenting with dyspnea during the spring months. All these findings could be taken into account when designing future research on the pathogenesis of the disease, and could guide decision-making for the prevention or early diagnosis of suspected cases at high risk.

There is a need for further large-scale prospective studies incorporating various hematological parameters and atmospheric pressure to explain the relationship between PE development and seasonal changes.

ETHICAL DECLARATIONS

Ethics Committee Approval

Our study was approved by the Afyonkarahisar Health Sciences University Non-interventional Clinical Researches Ethics Committee (Date: 03.05.2019, Decision No: 2019/7-173).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Arseven O, Bayındır Ü, Yıldırım N, et al. Venöz Tromboembolizm. Arseven O (ed). Temel akciğer sağlığı ve hastalıkları. 2. İstanbul: Nobel Tıp Kitabevleri; 2015.
2. Guidelines on the diagnosis and management of acute pulmonary embolism. Task force on pulmonary embolism, European Society of Cardiology. *Eur Heart J.* 2000;21(16):1301-1336. doi:10.1053/euhj.2000.2250
3. Van Belle A, Buller HR, Huisman MV, et al. Effectiveness of managing suspected pulmonary embolism using an algorithm combining clinical probability, D-dimer testing, and computed tomography. *JAMA.* 2006; 295(2):172-179. doi:10.1001/jama.295.2.172
4. Türk Toraks Derneği; Pulmoner tromboembolizm tanı ve tedavi uzlaşı raporu. 2015.
5. Meral M, Mirici A, Aslan S, et al. Barometric pressure andthe incidence of pulmonary embolism. *Chest.* 2005;128(4):2190-2194. doi:10.1378/chest.128.4.2190
6. Perincek G. Pulmoner Tromboemboli sıklığı ve mortalitesinin meteorolojik parametreler ile ilişkisi. Uzmanlık Tezi. Edirne: Trakya Üniversitesi, 2009.
7. Oztuna F, Ozsu S, Topbas M, Bülbül Y, Koşucu P, Ozlü T. Meteorological parameters and seasonal variations in pulmonary thromboembolism. *Am J Emerg Med.* 2008;26(9):1035-1041. doi:10.1016/j.ajem.2007.12.010
8. Anar C, Inal T, Erol S, et al. Are meteorological parameters a risk factor for pulmonary embolism? A retrospective analysis of 530 patients. *Balkan Med J.* 2015;32(3):279-284. doi:10.5152/balkanmedj.2015.15686
9. Ercan I, Coskun F, Cangur S, et al. An analysis of patients diagnosed with pulmonary embolism in terms of clinical and meteorological data. *Saudi Med J.* 2006;27(4):555-557.
10. Manfredini R, Gallerani M, Boari B, Salmi R, Mehta RH. Seasonal variation in onset of pulmonary embolism is independent of patients' underlying risk comorbid conditions. *Clin Appl Thromb Hemost.* 2004; 10(1):39-43. doi:10.1177/107602960401000106
11. Gallerani M, Boari B, Smolensky MH, et al. Seasonal variation in occurrence of pulmonary embolism: analysis of the database of the Emilia-Romagna region, Italy. *Chronobiol Int.* 2007;24(1):143-160. doi: 10.1080/07420520601139755
12. Guijarro R, Trujillo-Santos J, Bernal-Lopez MR, et al. Trend and seasonality in hospitalizations for pulmonary embolism: a time-series analysis. *J Thromb Haemost.* 2015;13(1):23-30. doi:10.1111/jth.12772

13. Stein PD, Kayali F, Beemath A, et al. Mortality from acute pulmonary embolism according to season. *Chest*. 2005;128(5):3156-3158. doi:10.1378/chest.128.5.3156

14. Masotti L, Ceccarelli E, Forconi S, Cappelli R. Seasonal variations of pulmonary embolism in hospitalized patients. *Respir Med*. 2005;99(11):1469-1473. doi:10.1016/j.rmed.2005.04.006

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1749167

Anatolian Curr Med J. 2025;7(5):670-677

Spiritual care needs and influencing factors among inflammatory bowel disease patients

Kamile Kırca[1], Elif Sözeri Öztürk[2], Şenay Uçan[3]

[1]*Department of Nursing, Faculty of Health Sciences, Kırıkkale University, Kırıkkale, Turkiye*

[2]*Department of Nursing, Faculty of Nursing, Gazi University, Ankara, Turkiye*

[3]*Division of Gastroenterology, Department of Internal Medicine, Faculty of Medicine, Gazi University, Ankara, Turkiye*

Cite this article as: Kırca K, Sözeri Öztürk E, Uçan Ş. Spiritual care needs and influencing factors among inflammatory bowel disease patients. *Anatolian Curr Med J.* 2025;7(5):670-677.

Received: 25.07.2025 ◆ **Accepted**: 06.09.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: This study aimed to determine the spiritual care needs of individuals diagnosed with inflammatory bowel disease (IBD) and to examine the relationship with some sociodemographic and clinical characteristics.

Methods: This descriptive and correlational study was conducted with individuals diagnosed with IBD (n=203). Data were collected using the "Descriptive Information Form," "Spiritual Care Needs Inventory," "Simple Clinical Colitis Activity Index," and " Harvey-Bradshaw Index." Descriptive statistics, t-test for independent groups, one-way analysis of variance, and Pearson correlation analysis were used to analyze the data.

Results: The mean total spiritual care need score of the participants was found to be 59.38±20.00. According to the sub-dimensions, the mean score of the "meaning and hope" sub-dimension was 35.56±13.45, and the mean score of the "caring and respecting" sub-dimension was 23.82±8.17. The "meaning and hope" sub-dimension scores of single and employed individuals were significantly higher than those of married and unemployed individuals ($p<0.05$). In addition, total spiritual care needs scores of individuals who did not use medication for IBD were found to be higher than those who used medication ($p<0.05$). A significant positive correlation was found between the Harvey-Bradshaw Index and the "caring and respecting" sub-dimension ($p<0.05$).

Conclusion: It was determined that the spiritual care needs of individuals diagnosed with IBD were generally above average. The increase in disease activity especially increases individuals' expectations for attention, being valued, and meaning in life. These findings suggest that nurses should be physically, spiritually, and psychosocially sensitive.

Keywords: Inflammatory bowel disease, spiritual care, spiritual need, chronic disease, nursing

INTRODUCTION

Inflammatory bowel diseases (IBD) are lifelong diseases characterized by chronic inflammation leading to progressive and irreversible destruction of the gastrointestinal tract.[1] IBDs negatively affect the quality of life of patients in many ways. Chronicisation of the disease, physical and psychosocial effects related to the disease, and treatment are the main reasons for the decrease in patients' quality of life.[2] The main physical symptoms affecting the quality of life of patients can be listed as abdominal pain, chronic and recurrent diarrhea attacks, sudden need for a toilet, nausea, vomiting, weight loss, anorexia, fatigue, etc.[3] Apart from the physical symptoms of the disease, psychosocial effects such as anxiety, depression, and social isolation also affect patients negatively.[4]

In diseases that require long-term treatment and in situations that reduce the quality of life, individuals' physical, emotional, spiritual and social needs as well as spiritual needs increase.[5] In addition, every patient with severe chronic illness needs basic spiritual care, including compassionate treatment.[6] However, spiritual and psychosocial needs are more abstract and complex than physical and difficult to measure.[5] The need for spiritual care refers to people's inner and dynamic dimension.[7] The fundamental element of holistic care is paying attention to the patient's spiritual needs.[8] Spiritual care can promote optimism and effective stress management, increase self-control, support self-care, and contribute to the regain of self-confidence and self-concept.[9] In a study examining the barriers to spiritual care for cancer patients in Turkiye, it was determined that nurses lacked knowledge about spiritual care, did not have enough time to provide this care, and there were insufficient private areas where they could discuss spiritual issues with the patient.[10]

Healthcare professionals recognize that coping with a chronic disease such as IBD can be extremely challenging and can have a significant negative impact on the patient's quality of life. These patients regularly worry about the many uncertainties associated with their disease, such as when

Corresponding Author: Kamile Kırca, kamilekirca@kku.edu.tr

the next flare-up will occur, whether they will be able to control the urge to defecate during social events or at work, whether current medications will wear off, and whether surgery will be required. These worries also affect their interactions with family members, children, colleagues, and friends, contributing to increased difficulties in coping with their illness and mood disorders.[11] A study conducted with individuals with chronic disease determined that increasing spiritual well-being also increases compliance with chronic disease.[12] In addition, in a study, it was stated that the spiritual needs of cancer patients are related to fatigue, depression, social support, and other factors.[13]

Although the physical and psychosocial impacts of IBD have been well documented, studies examining the spiritual care needs of these patients are limited. Addressing spiritual needs is a fundamental aspect of holistic care, which can enhance coping strategies, foster self-confidence, and improve overall quality of life. Understanding these needs is essential for healthcare professionals to provide patient-centered care that addresses not only physical symptoms but also emotional and spiritual well-being. Investigating the spiritual care needs of IBD patients therefore fills an important gap in the literature and contributes valuable insights to holistic care practices. Therefore, it is essential to meet the spiritual care needs of patients. This study aimed to determine the level of spiritual care needs of IBD patients to promote holistic care, improve their quality of life, and investigate the relationship with variables such as sociodemographic and disease/treatment characteristics.

METHODS

Ethics

Before starting the study, permission was obtained from the Gazi University Ethics Committee (Date: 30.04.2024, Decision No: 2024-827), and institutional permission was obtained. After receiving detailed information about the study, the patients provided written and verbal informed consent, voluntarily agreeing to participate. The research process was conducted in accordance with ethical principles, ensuring the confidentiality and rights of the participants, and the data were used solely for research purposes. The study was conducted according to the ethical principles outlined by the World Medical Association's Declaration of Helsinki.

Study Design and Setting

This study is a descriptive and correlational study. The study population was conducted in Gazi University Gastroenterology Department IBD Outpatient Clinic between June and November 2024. This clinic serves an average of 500 patients 5 days a week. Patients receive treatment during flare-ups of their disease and routine follow-ups during remission periods. Additionally, procedures for patients using anti-TNF and biological agents are coordinated with the infection department, and their treatments are scheduled. All patients receive education and psychological support regarding their disease, medications, and diet. Patients with rheumatological, dermatological, and autoimmune diseases are also monitored in coordination with the relevant departments. Depending on the patient's condition, support is also provided by a psychiatrist for patients with conditions such as depression and anxiety.

Study Population

In this study, convenience sampling was used as the sampling method. Participants were invited to the study during their routine outpatient visits, and those who volunteered were included in the study. Patients diagnosed with IBD who met the inclusion criteria, volunteered to participate in the study, and obtained written and verbal consent were included in the study sample. The inclusion criteria for the study sample were patients over 18 who could read and write, did not have communication difficulties, and volunteered to participate.

The study calculated the sample size at a 95% confidence level using the "G-power-3.1.9.2" program. As a result of the analysis calculated, the minimum sample size was 138 at α=0.05 level, based on a standardized effect size of 0.3 and a theoretical power of 0.95.[14] The study was completed with 203 patients.

Data Collection Tools

Data collection tools "Patient Introduction Form," "Simple Clinical Colitis Activity Index", "Harvey-Bradshaw Index" and "Spiritual Care Needs Inventory" were used to collect the research data.

Patient Introduction Form: It was created by the researchers in the light of literature information. It consists of questions related to sociodemographic data, health, disease, and treatment.[13,15,16]

Spiritual Care Needs Inventory (SCNI): The scale developed by Wu et al.[17] in 2016 provides information about the spiritual care needs of patients. A Turkish validity and reliability study was conducted by Günay İsmailoğlu et al.[5] in 2019. The scale differs from other scales because it can be used in all patients and patients with different religious beliefs, regardless of the reasons for hospitalization. The SCNI is a scale consisting of 21 items. The items in the scale include the potential spiritual care needs of the patients. Patients are asked to rate the necessity of the spiritual care needs in each item on a 5- point Likert scale. The evaluation is as follows: 1="not at all necessary", 2="not necessary", 3="does not matter", 4="necessary", 5="absolutely necessary". The average score that can be obtained from the scale varies between 21-105. The higher the mean total score on the scale, the more spiritual care the patient needs. The scale consists of 2 components: "meaning and hope" (1-12, 14) and "caring and respecting" (13, 15-21). The meaning and hope component includes expressions of spiritual well-being towards oneself, nature, and environmental factors; the caring and respect component includes expressions toward relationships with others. In the Turkish validity and reliability study of the scale, the general scale internal consistency Cronbach Alpha value was found to be 0.935.[5] In this study, the Cronbach Alpha value of the scale was determined as 0.957.

Simple Clinical Colitis Activity Index (SCCAI): A diagnostic tool and questionnaire used to assess the severity of symptoms

in people with ulcerative colitis. The index was created in 1998 and is still used to determine the severity of symptoms. Active disease scores five or higher.[18]

Harvey-Bradshaw Index (HBI): This index consists of questions that allow the patient to quickly categorize the severity of Crohn's disease and detect remission. Harvey and Bradshaw published the index in The Lancet in 1980 as a shorter and simpler alternative to the standard classification technique called the Crohn's Disease Activity Index. Patients answer five questions and are scored according to the severity of their symptoms. A score below 5 indicates remission.[19]

Data Collection

The study data were collected between June and November 2024 at a University Gastroenterology Department IBD Outpatient Clinic after obtaining the ethics committee's approval and institutional permission. Data were collected by the researchers through face-to-face interviews. Participants were invited to participate in the study during routine outpatient visits, and those who agreed to participate were taken to a quiet and appropriate room to ensure privacy and comfort. After providing detailed information about the study and obtaining written and verbal consent, the data collection process was initiated. The forms used were, in order, the "Patient Information Form," the "Simple Clinical Colitis Activity Index," the "Harvey-Bradshaw Index," and the "Spiritual Care Needs Inventory." Researchers only provided assistance when questions were unclear, refraining from any guidance that might influence responses. The entire application process took an average of 15 minutes per participant.

Statistical Analysis

The data obtained in the study were analyzed using SPSS (Statistical Package for Social Sciences) for Windows 25.0 and DataBeeg 1.0 software. Descriptive statistical methods (number, percentage, min-max values, mean, standard deviation) were used to evaluate the data. The data used were tested for conformity to normal distribution. Q-Q Plot drawing can analyze Compliance with a normal distribution.[20] In addition, the normal distribution of the data used depends on the skewness and kurtosis values being between ±3.[21] In the comparison of quantitative data in normally distributed data, an independent t-test was applied for the comparison of two independent groups, a one-way analysis of variance was applied for the comparison of more than two independent groups, and Bonferroni was used to find the group that made a difference in case of a difference. Pearson correlation was applied to test the relationship between two numerical variables.

RESULTS

The mean age of the patients who participated in the study was 41.79±14.51 years. 51.7% of the patients were female, 63.5% were married and 62.6% had children. The rate of primary school graduates was 40.9%, and the rate of employees was 66.0%. 56.7% of the patients reported their income level as "income equal to expenses." 26.1% of the patients were smokers, and 10.3% were alcohol users. 53.7% of the patients were diagnosed with Crohn's disease, and the proportion of patients with a diagnosis of IBD of 11 years or more was 30.5%. 45.3% of the patients were hospitalized due to IBD, and 19.2% underwent surgery due to IBD. It was determined that 61.1% of the patients had other chronic diseases/disorders in addition to IBD, and 55.2% of them were taking medication for these diseases/disorders. Harvey Bradshaw Activity Index in Crohn's disease patients: 3.88±2.61, and in patients with ulcerative colitis, the Ulcerative Colitis Simple Clinical Colitis Activity Index: 1.44±2.07 (**Table 1**).

The mean scores of the "meaning and hope" sub-dimension of the SCNI were 35.56±13.45, the mean scores of the "caring and respecting" sub-dimension were 23.82±8.17, and the mean total score of the scale was 59.38±20.00 (**Table 2**).

The comparison of the mean scores of the SCNI sub-dimension and total scores according to the characteristics of the patients participating in the study is given in **Table 3**. When the statistically significant differences were examined, it was determined that the mean scores of the "meaning and hope" sub-dimension of single patients were higher than those of married patients ($p<0.05$). It was also found that the mean scores of the "meaning and hope" sub-dimension of the employees were statistically significantly higher than the non-employees ($p<0.05$). The mean total scores of the SCNI of those who did not use medication for IBD were significantly higher than those who did not use medication ($p<0.05$).

Pearson correlation was applied to test the relationship between some characteristics of the patients and the mean scores of the SCNI sub-dimension and total scores. As a result, it was determined that there was a significant positive correlation between the mean scores of the "caring and respecting" sub-dimension and Harvey-Bradshaw Index scores ($p<0.05$). The relationships between other characteristics were not statistically significant ($p>0.05$) (**Table 4**).

DISCUSSION

The lack of standard clinical practice guidelines for addressing spirituality in adults with serious health problems reveals an essential deficiency in this field. Studies have shown that ignoring spiritual needs can lead to unresolved spiritual and physical pain, as well as avoidable distress and suffering.[22,23] This situation shows how vital spiritual care is, especially for individuals struggling with chronic and life-threatening diseases. It is seen that previous research has primarily focused on cancer patients, cancer survivors, or those at the end of their lives.[13,24-26] However, in the existing literature, studies specifically addressing the spiritual care needs of individuals with IBD are limited.[16,27] However, it is known that this patient group also experiences various difficulties at both physical, psychosocial, and spiritual levels due to chronic symptoms that negatively affect quality of life. In this respect, this study, which aims to evaluate the spiritual needs of IBD patients, has a unique quality that fills the existing gap in the literature and contributes to the understanding of holistic care.

In our study, it was determined that the spiritual care needs of individuals diagnosed with IBD were above average. When

Table 1. Distribution of the patients participating in the study according to their characteristics (n=203)

	n	%
Gender		
Female	105	51.7
Male	98	48.3
Marital status		
Single	74	36.5
Married	129	63.5
Child presence		
There is	127	62.6
None	76	37.4
Education status		
Primary and secondary school	43	21.2
High School	83	40.9
Undergraduate and postgraduate	77	37.9
Employment status		
Yes	134	66.0
No	69	34.0
Income level		
Income less than expenditure	50	24.6
Income equals expenditure	115	56.7
Income more than expenditure	38	18.7
Smoking status		
Yes	53	26.1
No	150	73.9
Alcohol use status		
Yes	21	10.3
No	182	89.7
Medical diagnosis		
Ulcerative colitis	94	46.3
Crohn's disease	109	53.7
Time to diagnosis of IBD		
0-5 years	107	52.7
6 years and over	96	47,3
Hospitalization due to IBD		
Yes	92	45.3
No	111	54.7
Operation due to IBD		
Yes	39	19.2
No	164	80.8
Presence of chronic disease		
Yes	124	61.1
No	79	38.9
Medication use due to chronic disease		
Yes	112	55.2
No	91	44.8
Age: 41.79±14.51 (minimum=18.00, maximum=83.00)		
Harvey-Bradshaw Index: 3.88±2.61		
Simple Clinical Colitis Activity Index: 1.44±2.07		

IBD: Inflammatory bowel disease

Table 2. Patients' mean scores of the Spiritual Care Needs Inventory (SCNI) subscale and total scores (n=203)

	Mean±SD	Minimum	Maximum
Meaning and hope	35.56±13.45	13.00	65.00
Caring and respecting	23.82±8.17	8.00	40.00
Total score	59.38±20.00	21.00	105.00

SD: Standard deviation

Table 3. Comparison of the mean scores of the Spiritual Care Needs Inventory (SCNI) subscale and total scores according to the characteristics of the patients (n=203)

Characteristics		Meaning and hope	Caring and respecting	SCNI total score
		Mean±SD	Mean±SD	Mean±SD
Gender	Female	35.05±12.69	24.31± 8.44	59.36± 19.48
	Male	35.50± 13.69	23.30± 7.90	58.80± 20.11
	Test value	-0.244**	0.886**	0.204**
	p	0.807	0.377	0.839
Marital status	Single	36.36± 13.01	23.91± 8.16	60.26± 20.01
	Married	33.36± 13.27	23.68± 8.26	57.04± 19.22
	Test value	2.166**	0.194**	1.120**
	p	0.031*	0.847	0.264
Child presence	Yes	35.75±13.19	23.83±8.14	59.58±19.88
	No	34.46±13.13	23.80±8.30	58.26±19.60
	Test value	0.674**	0.027**	0.460**
	p	0.501	0.979	0.646
Education status	Primary and secondary school	37.23±13.47	24.79±8.43	62.02±20.37
	High school	34.51±13.18	23.39±8.60	57.89±20.33
	Undergraduate and postgraduate	34.99±13.01	23.75±7.61	58.74±18.81
	Test value	0.6434***	0.420***	0.638***
	p	0.531	0.657	0.530
Employment status	Yes	36.04±13.67	23.84±8.70	59.87±20.89
	No	33.77±12.04	23.80±7.12	57.57±17.32
	Test value	2.168**	0.032**	0.788**
	p	0.025*	0.975	0.431
Income level	Income less than expenditure	36.68±12.56	24.04±8.19	60.72±19.40
	Income equals expenditure	35.47±12.98	23.90±8.02	59.37±19.51
	Income more than expenditure	32.79±14.39	23.29±8.81	56.08±21.03
	Test value	0.976***	0.103***	0.622***
	p	0.378	0.902	0.538
Smoking status	Yes	34.26±11.70	22.85±8.46	57.11±18.77
	No	35.62±13.65	24.17±8.08	59.79±20.08
	Test value	-0.644**	-1.008**	-0.847**
	p	0.520	0.314	0.398

The table continues

Table 3. Comparison of the mean scores of the Spiritual Care Needs Inventory (SCNI) subscale and total scores according to the characteristics of the patients (n=203) (The table continues)

Alcohol use status	Yes	35.90±14.16	24.67±8.29	60.57±20.96
	No	35.19±13.07	23.73±8.18	58.92±19.65
	Test value	0.235**	0.499**	0.363**
	p	0.815	0.619	0.717
Medical diagnosis	Ulcerative colitis	34.94±13.03	23.41±7.74	58.35±19.21
	Crohn's disease	35.55±13.31	24.17±8.56	59.72±20.25
	Test value	-0.331**	-0.659**	-0.494**
	p	0.741	0.511	0.622
Time to diagnosis of IBD	0-5 years	35.53±13.92	23.76±8.30	59.29±20.56
	6 years and over	34.97±12.30	23.90±8.08	58.86±18.88
	Test value	0.304**	-0.120**	0.153**
	p	0.761	0.904	0.879
Medication use status	Yes	35.27±13.21	23.75±8.11	59.02±19.75
	No	35.25±11.62	27.50±11.79	62.75±21.75
	Test value	0.002**	-0.908**	-2.271**
	p	0.998	0.365	0.025*
Hospitalization due to IBD	Yes	35.40±12.00	23.52±8.28	58.92±18.67
	No	35.15±14.08	24.07±8.12	59.23±20.67
	Test value	0.134**	-0.476**	-0.108**
	p	0.894	0.634	0.914
Operation due to IBD	Yes	36.77±14.12	23.33±8.87	60.10±22.19
	No	34.91±12.93	23.94±8.03	58.85±19.18
	Test value	0.793**	-0.415**	0.356**
	p	0.428	0.679	0.722
Presence of chronic disease	Yes	35.50±12.87	23.87±8.21	59.37±19.37
	No	34.90±13.66	23.75±8.18	58.65±20.42
	Test value	0.317**	0.105**	0.255**
	p	0.752	0.916	0.799
Medication use due to chronic disease	Yes	35.61±13.16	23.97±8.43	59.58±20.06
	No	34.85±13.21	23.64±7.90	58.48±19.43
	Test value	0.409**	0.290**	0.393**
	p	0.683	0.772	0.695

SCNI: Spiritual Care Needs Inventory, IBD: Inflammatory bowel disease, SD: Standard deviation, *p<0.05, **Independent t-test, ***One-way analysis of variance

Table 4. The relationship between some characteristics of the patients and the mean scores of the subscales and total scores of the Spiritual Care Needs Inventory

	Age		Harvey-Bradshaw Index		Simple Clinical Colitis Activity Index	
	r	p	r	p	r	p
Meaning and hope	0.045	0.521	-0.046	0.515	-0.054	0.442
Caring and respecting	0.023	0.746	0.196	0.047*	0.048	0.497
Total score	0.040	0.574	0.000	0.996	-0.016	0.818

r: Pearson correlation coefficient, *p<0.05

evaluated in general, it can be said that the spiritual care needs of the patients are significant and at a level that should not be ignored. This finding parallels the results of studies conducted with individuals with different chronic conditions in the literature.[6,13,28] For individuals with serious illnesses, spirituality stands out as an essential source of support in the process of coping with life. Studies show that most individuals (71-99%) find spirituality meaningful, and more than half (50-96%) demand spiritual care within the scope of health services. However, despite this high demand, it is seen that spiritual care needs are not sufficiently met in clinical practices. It is stated that the rates of patients who report that the health system does not meet their spiritual needs vary between 49% and 91%.[22,29] This situation shows that spiritual care is still secondary in clinical practice, and the holistic care approach has not been fully realized. It is thought that the main reasons for the high need for spiritual care in individuals diagnosed with IBD may be related to the chronic, unpredictable nature of the disease, which directly affects the quality of life. At the same time, psychosocial burden and social stigmatization increase the search for meaning, the need for hope, and the search for inner support.

Long-term hospitalization of patients treated in internal medicine clinics is an important factor that increases their spiritual care needs.[30] In particular, individuals with chronic diseases or faced with conditions such as IBD that radically affect life may enter a deep spiritual questioning process over time. In this process, they may perceive the disease as divine punishment, experience a sense of injustice, and lose hope in prayer and belief systems. Such spiritual distress may damage not only the psychological but also the spiritual well-being of the person.[27] However, data on the spiritual care needs of individuals diagnosed with IBD are limited. However, determining these needs is essential in developing patient-centered care practices and planning appropriate support services. In previous studies, it was reported that adult individuals stated that spirituality helped them to reinterpret the disease process, gain a new perspective on their lives, and experience inner relief during challenging periods.[31] In a study conducted by Starnella et al.[16] with individuals with IBD (n=103), it was determined that patients developed their spirituality by maintaining their lives by their core values (47.7%), staying in contact with nature (30.5%), praying (15.8%), engaging in introspection and self-knowledge activities (14.7%), volunteering (12.6%) and doing extreme sports (10.5%).

Various studies have suggested that spirituality may positively affect psychosocial outcomes in individuals with chronic physical illnesses.[31,32] Although the exact mechanisms through which this effect occurs cannot be clearly explained, it is thought that spirituality contributes to individuals coping with the disease more harmoniously, which may result in a decrease in anxiety and depression levels, an increase in quality of life, and alleviation of physical symptoms. In a meta-synthesis study conducted by Hodge and Horvath,[33] 11 qualitative studies were analyzed on the spiritual needs

expressed by individuals with non-fatal health problems in healthcare settings. The main themes that emerged from this analysis were summarised as the search for meaning, purpose, and hope in life; emotional needs related to a relationship with God (e.g., guilt or questioning); the need to engage in spiritual practices such as prayer, meditation, and scripture reading; sensitivities towards the fulfillment of religious obligations (e.g. special dietary practices); the desire for strong social bonds with family members and clergy; and finally the desire for an empathetic, supportive relationship with health professionals (e.g. compassionate approach and advocacy).[33] These findings show that spiritual care is a multidimensional need that can contribute not only to the physical aspect of the disease but also to the individual's holistic well-being.

In our study, the sub-dimension of meaning and hope indicates a need slightly above the medium level. It shows that patients have a measured but significant need to search for inner meaning, a sense of hope, and harmony with nature and the environment. It shows that patients are more focused on personal belief, meaning of life, hope, and inner integrity and need more support. The meaning and hope component also includes spiritual well-being towards oneself, nature, and environmental factors.[5] In the process of coping with chronic diseases, individuals may face vital crises such as loss of status and role, fear of death, insufficient social support, loneliness, pain, and hopelessness about the future. In such periods, patients' spiritual needs, such as feeling safe, adding meaning to their lives, receiving love, and belonging to a place, come to the fore.[34] In this context, it is not surprising that this dimension emerges as an essential need in individuals diagnosed with IBD.

In our study, the moderate level of need in the "caring and respect" sub-dimension of the spiritual care scale shows that patients have a confident expectation of being respected, feeling valuable, and receiving social support from the people around them. This sub-dimension points to the importance of a caring approach based on empathy, respect, and sensitivity to individual values, especially in relationships with others.[5] The understanding and caring approach shown in patient-physician and patient-nurse interactions makes patients feel more secure and understood. The fact that patients want to be seen and cared for both physically and spiritually reveals the need to address them with a more holistic care approach. This finding coincides with the study conducted by Zumstein-Shaha et al.[7] with cancer patients. In this study, nurses stated that individuals living with chronic illnesses were not only physiological but also questioned the meaning of life and needed to give their lives a purpose. While some participants saw overcoming chemotherapy as a life goal, others started spending more time with family members or their pets as their primary goal. These examples reveal that individuals need to be supported biologically, emotionally, socially, and spiritually during the disease process. In addition, some patients reported that maintaining their ties with their faith communities was essential to spiritual support. These individuals ask their faith communities to pray for them, and nurses respect the patient's spirituality by integrating this request into the care process.[7] All these findings show that including the individual's spiritual, social, and emotional aspects in the care process can increase patient satisfaction and well-being.

In our study, significant differences were observed in the spiritual care needs of individuals diagnosed with IBD according to their demographic characteristics. In particular, the higher scores of single individuals in the "meaning and hope" sub-dimension compared to married individuals suggest that this group may have a more pronounced need for loneliness, lack of emotional support and sense of belonging. Lack of spousal or family support may cause individuals to turn more to internal resources and spiritual support systems in coping with the disease. Indeed, Büssing et al.[35] and Höcker et al.[36] similarly stated that spiritual coping mechanisms are activated more intensively in individuals who lack social support. In addition, in a study conducted by Shi et al.[13] with cancer patients, it was reported that the spiritual care needs of widowed and divorced individuals were significantly higher than those of married individuals. These findings show that family structure and social support networks are among the critical factors affecting the level of spiritual needs of the individual and point to the importance of considering the social situation in patient-centered care.

Similarly, it is noteworthy that the scores of working individuals in this sub-dimension are significantly higher than those of non-working individuals. Working life may increase the need to question more about the meaning of life, the desire to achieve success, or the need to cope with stress, and this may trigger the need for spiritual support. Another important finding was that the total spiritual care needs scores of individuals who did not use medication for IBD were higher than those who used medication. This suggests that access to medical treatment or perception of treatment may affect spiritual needs in the management of the disease. Individuals who do not use medication may tend to turn more to spiritual resources to cope with the disease. All these findings reveal that individual and clinical characteristics may affect spiritual care needs and it is essential to consider these variables in care planning.

Our study found a significant positive correlation between the "caring and respecting" sub- dimension and Harvey-Bradshaw Index scores. This suggests that as the symptoms of IBD become more severe, patients expect more attention and understanding from their social environment and health professionals. This reveals that spiritual care needs are internal and have a relational/psychosocial dimension. This finding reveals that individuals desire to be cared for and seen more in coping with the disease, and requires that caring professionals respond to this need with a sensitive approach. The fact that no significant relationship was found between the need for spiritual care and other variables suggests that spiritual needs are shaped more by the course of the disease and individual coping capacity. These results emphasize the importance of planning spiritual care in chronic diseases in a manner sensitive to disease activity.

Limitations

This study has some limitations. First, since the study was conducted at a single center, the findings cannot be directly generalized to all IBD patients; however, they are indicative for patient groups with similar characteristics. In addition, the fact that spirituality is a concept that can show individual and cultural differences may make it difficult to reach the same results in similar studies to be conducted in different social structures. Finally, the study did not assess whether the participants had received any professional spiritual care services before; this shows that an important variable that may affect the spiritual need levels of individuals is ignored. In line with these limitations, obtaining more in-depth information with longitudinal designs and qualitative data collection methods in future studies is recommended.

CONCLUSION

This study revealed that individuals with IBD have significant spiritual care needs, and these needs should not be ignored in clinical care processes. The findings indicate that nurses should be sensitive not only to physical symptoms but also to the spiritual and psychosocial needs of individuals. Especially in the axis of "meaning and hope," creating an environment where individuals can express their thoughts about the meaning of life, and in the dimension of "caring and respect," protecting privacy, respecting individual beliefs, and encouraging active listening should be among the basic components of patient-centered spiritual care. Coping skills play a key role in the self-management of chronic diseases, and nurses' recognition and support of patients' spiritual coping resources can strengthen self-management processes. In this context, making nurses more aware and equipped to assess spiritual needs can potentially improve the quality of care. However, the limited literature on spirituality in individuals with chronic diseases, especially in individuals diagnosed with IBD, and the lack of comparative studies with healthy individuals draw attention. In this respect, this study can provide a basis for future longitudinal and multi-method studies.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was conducted with the permission of Gazi University Ethics Committee (Date: 30.04.2024, Decision No: 2024-827).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The author declares no conflicts of interest.

Financial Disclosure

The author declares that this study received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

Acknowledgements

The authors acknowledge and thank all patients who kindly completed the survey.

REFERENCES

1. Le Berre C, Danese S, Peyrin-Biroulet L. Can we change the natural course of inflammatory bowel disease? *Ther Adv Gastroenterol*. 2023;16: 17562848231163118. doi:10.1177/17562848231163118
2. Norouzkhani N, Faramarzi M, Ghodousi Moghadam S, et al. Identification of the informational and supportive needs of patients diagnosed with inflammatory bowel disease: a scoping review. *Front Psychol*. 2023;14:1055449. doi:10.3389/fpsyg.2023.1055449
3. Schoefs E, Vermeire S, Ferrante M, et al. What are the unmet needs and most relevant treatment outcomes according to patients with inflammatory bowel disease? A qualitative patient preference study. *J Crohns Colitis*. 2023;17(3):379-388. doi:10.1093/ecco-jcc/jjac145
4. Alarhayem A, Achebe E, Logue AJ. Psychosocial support of the inflammatory bowel disease patient. *Surg Clin North Am*. 2015;95(6): 1281-1293. doi:10.1016/j.suc.2015.08.005
5. Günay İsmailoğlu E, Özdemir H, Erol A, Zaybak A. Spiritüel bakım gereksinimleri ölçeği Türkçe formunun geçerlik ve güvenirliği. *Dokuz Eylül Üniversitesi Hemşirelik Fakültesi Elektronik Dergisi*. 2019;12(4): 255-263.
6. Klimasiński M, Baum E, Praczyk J, et al. Spiritual distress and spiritual needs of chronically ill patients in Poland: a cross-sectional study. *Int J Environ Res Public Health*. 2022;19(9):5512. doi:10.3390/ijerph19095512
7. Zumstein-Shaha M, Ferrell B, Economou D. Nurses' response to spiritual needs of cancer patients. *Eur J Oncol Nurs*. 2020;48:101792. doi: 10.1016/j.ejon.2020.101792
8. Baldacchino D. Spiritual care education of health care professionals. *MDPI*. 2015;6(2):594-613. doi:10.3390/rel6020594
9. Ramezani M, Ahmadi F, Mohammadi E, Kazemnejad A. Spiritual care in nursing: a concept analysis. *Int Nurs Rev*. 2014;61(2):211-219. doi:10. 1111/inr.12099
10. Uslu Şahan F, Terzioğlu F. Hemşirelerin kanser hastalarına yönelik manevi bakım uygulamaları ve karşılaştıkları engeller. *Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi*. 2021;9(4):340-349. doi:10.37989/ gumussagbil.753529
11. Jedel S, Hankin V, Voigt RM, Keshavarzian A. Addressing the mind, body, and spirit in a gastrointestinal practice for inflammatory bowel disease patients. *Clin Gastroenterol Hepatol*. 2012;10(3):244-246. doi:10. 1016/j.cgh.2011.11.028
12. Kütmeç Yılmaz C, Kara FŞ. The effect of spiritual well-being on adaptation to chronic illness among people with chronic illnesses. *Perspect Psychiatr Care*. 2021;57(1):318-325. doi:10.1111/ppc.12566
13. Shi X, Wang F, Xue L, et al. Current status and influencing factors of spiritual needs of patients with advanced cancer: a cross-sectional study. *BMC Nurs*. 2023;22(1):131. doi:10.1186/s12912-023-01306-9
14. Cohen J. Statistical power analysis for the behavioral sciences. 2nd ed. Erlbaum. 1988.
15. Alvestad L, Jelsness-Jørgensen LP, Goll R, et al. Health-related quality of life in inflammatory bowel disease: a comparison of patients receiving nurse-led versus conventional follow-up care. *BMC Health Serv Res*. 2022;22:1602. doi:10.1186/s12913-022-08985-1
16. Starnella G, Rosato I, Canova C, et al. N47 Spiritual needs of patients with IBD: preliminary results of a pilot study. *J Crohns Colitis*. 2024; 18(Suppl 1):i2256. doi:10.1093/ecco-jcc/jjad212.1419
17. Wu LF, Koo M, Liao YC, Chen YM, Yeh DC. Development and validation of the spiritual care need inventory for acute care hospital patients in Taiwan. *Clin Nurs Stud*. 2016;25(6):590-606. doi:10.1177/ 1054773815579609
18. Walmsley RS, Ayres RC, Pounder RE, Allan RN. A simple clinical colitis activity index. *Gut*. 1998;43(1):29-32. doi:10.1136/gut.43.1.29
19. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet*. 1980;1(8167):514. doi:10.1016/S0140-6736(80)92767-1

20. Chan YH. Biostatistics 101: data presentation. *Singapore Med J*. 2003; 44(6):280-285.
21. Shao AT. Marketing research: an aid to decision making. South-Western/Thomson Learning. 2002.
22. Balboni TA, VanderWeele TJ, Doan-Soares SD, et al. Spirituality in serious illness and health. *JAMA*. 2022;328(2):184-197. doi:10.1001/jama.2022.11086
23. Varner-Perez SE, Comer AR, Fitchett G. Time to develop guidelines for spiritual care in serious illness. *Palliat Med Rep*. 2024;5(1):396-398. doi:10.1089/pmr.2024.0035
24. Acar G, Sağkal Midilli T. Examination of spiritual care needs of oncology patients and spiritual care competencies of oncology nurses. *İzmir Katip Çelebi Üniversitesi Sağlık Bilimleri Fakültesi Dergisi*. 2023;8(3):917-923. doi:10.61399/ikcusbfd.1134673
25. Epstein-Peterson ZD, Sullivan AJ, Enzinger AC, et al. Examining forms of spiritual care provided in the advanced cancer setting. *Am J Hosp Palliat Care*. 2015;32(7):750-757. doi:10.1177/1049909114540318
26. Sherman AC, Merluzzi TV, Pustejovsky JE, et al. A meta-analytic review of religious or spiritual involvement and social health among cancer patients. *Cancer*. 2015;121(21):3779-3788. doi:10.1002/cncr.29352
27. Martinato M. Spiritual needs of patients with inflammatory bowel disease and their caregivers: a European cross-sectional nursing study. *J Crohns Colitis*. 2024;18(Suppl 1):i2262. doi:10.1093/ecco-jcc/jjad212.1437
28. Kırca K, Özveren H, Bağcı S. Evaluation of the spiritual care needs of patients with cancer in a chemotherapy unit in Turkey. *J Relig Health*. 2024;63:1457-1474. doi:10.1007/s10943-023-01876-7
29. The Lancet Regional Health–Europe. Time to integrate spiritual needs in health care. *Lancet Reg Health Eur*. 2023;28:100648. doi:10.1016/j.lanepe.2023.100648
30. Ercan F, Körpe G, Demir S. Bir üniversite hastanesinde yataklı servislerde çalışan hemşirelerin maneviyat ve manevi bakım algıları. *Gazi Tıp Dergisi*. 2018;29(1):17-22. doi:10.12996/gmj.2018.05
31. Cotton S, Kudel I, Roberts YH, et al. Spiritual well-being and mental health outcomes in adolescents with or without inflammatory bowel disease. *J Adolesc Health*. 2009;44(5):485-492. doi:10.1016/j.jadohealth.2008.09.013
32. Riley BB, Perna R, Tate DG, et al. Types of spiritual well-being among persons with chronic illness: their relation to various forms of quality of life. *Arch Phys Med Rehabil*. 1998;79(3):258-264. doi:10.1016/S0003-9993(98)90004-1
33. Hodge DR, Horvath VE. Spiritual needs in health care settings: a qualitative meta-synthesis of clients' perspectives. *Soc Work*. 2011;56(4):306-316. doi:10.1093/sw/56.4.306
34. Çelik Y, Karaca Sivrikaya S. Dahili kliniklerde çalışan hemşirelerin manevi bakım yaklaşımı. *Kocatepe Tıp Dergisi*. 2024;25(1):1-7. doi:10.18229/kocatepetip.1199183
35. Büssing A, Balzat HJ, Heusser P. Spiritual needs of patients with chronic pain diseases and cancer—validation of the Spiritual Needs Questionnaire. *Eur J Med Res*. 2010;15:266-273. doi:10.1186/2047-783X-15-6-266
36. Höcker A, Krüll A, Koch U, Mehnert A. Exploring spiritual needs and their associated factors in an urban sample of early and advanced cancer patients. *Eur J Cancer Care (Engl)*. 2014;23:786-794. doi:10.1111/ecc.12200

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1759424

Anatolian Curr Med J. 2025;7(5):678-686

Awareness of procedure specific pain management (PROSPECT) in pain management after total hip arthroplasty surgery among anesthesiology and reanimation physicians survey study

Dilek Yeniay, Ali Altınbaş

Department of Anesthesiology and Reanimation, Faculty of Medicine, Giresun University, Giresun, Turkiye

Cite this article as: Yeniay D, Altınbaş A. Awareness of procedure specific pain management (PROSPECT) in pain management after total hip arthroplasty surgery among anesthesiology and reanimation physicians survey study. *Anatolian Curr Med J.* 2025;7(5):678-686.

Received: 06.08.2025 ◆ **Accepted**: 06.09.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: Procedure-specific pain management (PROSPECT) provides clinicians with evidence-based recommendations for postoperative pain management. We aimed to evaluate Anesthesiology and Reanimation physicians' preferences for analgesia management after total hip arthroplasty (THA) surgery.

Methods: A questionnaire consisting of 19 questions including demographic data, analgesia methods used for THA surgery and whether PROSPECT recommendations were followed was prepared. This questionnaire was sent to the participants via e-mail.

Results: A total of 199 participants responded to the survey. 43.8% of the participants reported following PROSPECT recommendations. Only 6.5% of the participants did not prefer opioids for analgesia after THA surgery, while 14.1% reported that they routinely used opioids and the rest preferred them in combination with other techniques.

Conclusion: Since THA surgery is mostly performed in the elderly population with comorbid diseases, there is no consensus among anesthesiology and reanimation physicians about postoperative analgesia methods. We think that postoperative pain management trainings in THA surgery should be increased and PROSPECT recommendations should be disseminated in these trainings.

Keywords: Anesthesiology and reanimation, survey, PROSPECT, total hip arthroplasty, pain management

INTRODUCTION

The procedure-specific pain management (PROSPECT) working group is a global collaboration of surgeons and anesthesiologists formulating surgical procedure-specific recommendations for pain management after common operations.[1] PROSPECT recommendations are based on a procedure-specific systematic literature review of randomized controlled trials and systematic reviews. The methodology reports true clinical efficacy, balancing the invasiveness of analgesic interventions and the level of postoperative pain against efficacy and side effects, while setting general recommendations. In addition, attention is paid to early rehabilitation and mobilization.[2]

Total hip arthroplasty (THA) surgery is a common surgical procedure aimed at improving mobility and quality of life, especially in elderly patients.[3] The choice of postoperative analgesia is important because the patient group is especially elderly patients with comorbid diseases. Adequate analgesia with minimal side effects provides early postoperative mobility, optimal functional recovery, and reduced postoperative morbidity.[4,5] Despite being a common surgical procedure, there is no "gold standard" in the literature for the control of postoperative pain after THA surgery, and there is high variability in perioperative anesthetic and analgesic management.[6] This creates variability in clinicians' postoperative analgesia preferences.

The PROSPECT recommendations for pain management after THA surgery were first published in 2005 and then updated periodically with new evidence and the final guideline was published in 2021. In line with these updates, recommendations for analgesic interventions in pain management after THA surgery are listed in **Table 1** and non-recommended interventions are listed in **Table 2**.[7]

The aim of this study was to evaluate Anesthesiology and Reanimation physicians' pain management practices after THA and their awareness of PROSPECT recommendations during these practices.

METHODS

The study was approved by the Giresun Training and Research Hospital Ethics Committee (Date: 19.03.2025, Decision No: 19.03.2025/05). All procedures were carried out in accordance with the ethical rules and the principles

Corresponding Author: Dilek Yeniay, dgyeniay@gmail.com

Table 1. General recommendations for pain management in patients undergoing total hip arthroplasty
Before and during surgery
Preoperative exercise and education (grade A)
General or spinal anesthesia (grade A)
Paracetamol (grade A)
Non-steroidal anti-inflammatory drugs or cyclo-oxygenase-2 selective inhibitors (grade A)
Dexamethasone 8-10 mg i.v. (grade A)
Single shot fascia iliaca block or local infiltration analgesia (LIA) (grade D)
Intrathecal morphine 0.1 mg is acceptable if the patient received spinal anesthesia during surgery (grade D)
After surgery
Paracetamol (grade A)
Non-steroidal anti-inflammatory drugs or cyclo-oxygenase-2 selective inhibitors (grade A)
Rescue opioid (grade D)

of the Declaration of Helsinki.The study was a descriptive type study and data were collected with a questionnaire form created in electronic environment (Google forms). In order to reach the anesthesiology and reanimation physicians in Turkiye, the questionnaire form was sent to the members of the Turkish Society of Anesthesiology and Reanimation (TARD) via e-mail twice and dissemination was ensured. The questions were directed to the participants between 01.04.2025-01.06.2025. Detailed information about the purpose and nature of the questionnaire for the participants was stated in the introduction part of the questionnaire. The first 5 questions included demographic data and the next 13 questions included pain management data after THA surgery. The questionnaire form was administered online to actively working anesthesiology and reanimation physicians with the help of digital media. Within the specified date range, 199 participants responded to the questionnaire and the data of the participants who completed the questionnaire form completely were analyzed.

Survey Form;

Section 1 questions:

1. Age: a) 20-30 b) 31-40 c) 41-50 d) 51-60 e) over 60

2. Sex: a) Female b) Male

3. Title: a) Research Assistant b) Specialist Physician c) Assistant Professor d) Associate Professor e) Professor

4. Professional experience: a) 0-10 b) 11-20 c) 20 and above

5. Active organization: a) University Hospital b) Ministry of Health Training and Research Hospital c) State Hospital d) Private Hospital

Section 2 questions:

1. Do you recommend providing education about exercise and surgery during the preoperative period for pain control after total hip arthroplasty surgery?

a) Yes b) No

2. What is the method you most frequently use for pain control after total hip arthroplasty surgery?

a) General anesthesia b) Spinal anesthesia

Table 2. Analgesic interventions not recommended for pain management in patients undergoing total hip arthroplasty

	Intervention	**Reason not to recommend**
Before or during surgery	Carbohydrate loading	Limited procedure-specific evidence
	Day treatment	Limited procedure-specific evidence
	COX-2-selective inhibitor before or after incision	Limited procedure-specific evidence
	Gabapentinoids	Inconsistent evidence for single doses.
		Procedure-specific evidence for multiple perioperative doses, but extra side effects
	Ketamine	Limited procedure-specific evidence
	Lateral femoral cutaneous block	Limited procedure-specific evidence
	Anterior quadratus lumborum block	Limited procedure-specific evidence
	Femoral nerve block	Procedure-specific evidence, but side effects
	Lumbar plexus block	Procedure-specific evidence, but side effects
	LIA (local infiltration analgesia) is used in addition to local anesthetic drugs	Inconsistent procedure-specific evidence
	LIA infusion or repeated injections	Inconsistent procedure-specific evidence
	Epidural analgesia	Procedure-specific evidence, but side effects
After surgery	Tranexamic acid	Lack of procedure-specific evidence
	Partial load execution	Lack of procedure-specific evidence
	Topical fibrin glue	Lack of procedure-specific evidence
	TENS (transcutaneous electrical nerve stimulation)	Limited procedure-specific evidence
Surgical technique	Anterior approach and posterolateral approach	Inconsistent procedure-specific evidence
	Minimally invasive and traditional incision	Inconsistent procedure-specific evidence, increased risks

COX-2: Cyclooxygenase, LIA: Local infiltration analgesia, TENS: Transcutaneous Electrical Nerve Stimulation

3. Do you use intravenous paracetamol before or during surgery for pain control after total hip arthroplasty surgery?

a) Yes b) No

4. Do you use NSAIDs or COX-2 selective inhibitors preoperatively or intraoperatively for pain control after total hip arthroplasty surgery?

a) Yes b) No

5. Do you use a single dose of intravenous dexamethasone before or during surgery for pain control after total hip arthroplasty surgery?

a) Yes b) No

6. Do you add an intrathecal long-acting opioid (morphine 0.1 mg) to spinal anesthesia for pain control in patients who have undergone total hip arthroplasty surgery?

a) Yes b) No

7. Do you prefer patient-controlled epidural analgesia for pain control after total hip arthroplasty surgery?

a) Yes b) No

8. Do you use peripheral nerve blocks for pain control after total hip arthroplasty surgery? If so, what is your most preferred peripheral block? (You may select more than one option)

a) I do not use them

b) Femoral nerve block

c) Lumbar plexus block

d) Psoas compartment block

e) Fascia iliaca block

f) Lateral femoral cutaneous nerve block

g) Anterior quadratus lumborum block

h) Pericapsular nerve block (PENG)

9. Do you use local infiltration analgesia for pain control after total hip arthroplasty surgery?

a) Yes b) No

10. What type of intravenous analgesic do you most frequently administer in the postoperative period for pain control after total hip arthroplasty surgery?

a) Paracetamol

b) NSAID

c) Tramadol

d) Fentanyl

e) Pethidine

11. What is your frequency of opioid use in the postoperative period for pain control after total hip arthroplasty surgery?

a) I use it routinely

b) I use it in combination with regional techniques

c) I use it in combination with paracetamol/NSAIDs

d) I use it when I cannot provide adequate analgesia with other methods

e) I do not use opioids

12. Do you recommend analgesic adjuvants (such as tranexamic acid, partial weight-bearing walking, and TENS) for pain control in the postoperative period following total hip arthroplasty surgery?

a) Yes b) No

13. Do you follow the PROSPECT recommendations?

a) Yes b) No

Statistical Analysis

Statistical evaluation of the data was performed using the IBM Statistical Package for the Social Sciences (SPSS) 22.0 (Version 22.0. Armonk, NY: IBM Corp.) package program. Descriptive statistics of categorical variables were given as frequency and percentage.

RESULTS

A total of 199 anesthesiology and reanimation physicians participated in our survey. The demographic data of the participants (sex, age ranges, title, years of professional experience and active institution) are summarized in **Table 3**.

Table 3. Demographic data

n=199		**n (%)**
Sex	Female	113 (56.7%)
	Male	86 (43.2%)
Age(years)	20-30	17 (8.5%)
	31-40	89 (44.7%)
	41-50	59 (29.6%)
	51-60	29 (14.6%)
	Over 60 years old	5 (2.5%)
Title	Research Assistant	118 (59.3%)
	Specialist Physician	38 (19.1%)
	Assistant Professor	13 (6.5%)
	Associate Professor	19 (9.5%)
	Professor Doctor	11(5.5%)
Professional experience (years)	0 to 10	92 (46.2%)
	11 to 20	55 (27.6%)
	20 and above	52 (26.1%)
Active organization	University Hospital	48 (24.1%)
	Ministry of Health Training and Research Hospital	32 (16.1%)
	State Hospital	79 (39.7%)
	Private Hospital	40 (20.1%)

Categorized as *n and % columns

The general responses to the questions determined for PROSPECT recommendations are shown in **Table 4**. It was observed that 43.2% (n=86) of the participants followed

PROSPECT recommendations for pain control after THA surgery. It was observed that 74.4% of the physicians who participated in our survey recommended preoperative exercise and training. 92.9% (n=185) of the participants preferred spinal anesthesia for pain control after THA surgery, while only 7.1% (n=14) preferred general anesthesia. For pain control after THA surgery, 73.3% of the participants used paracetamol, 64.8% used NSAIDs or COX2 selective inhibitors, and 26.1% used a single dose of intravenous dexamethasone before and during surgery. The most preferred analgesic drug in the postoperative period was tramadol with 49.2% (**Figure 1**). In addition to spinal anesthesia, 33.7% of the participants used an intrathecal long-acting opioid (morphine 0.1 mg), 72.9% preferred patient-controlled epidural analgesia, and 18.6% preferred local infiltration analgesia (LIA). Only 7.1% of the participants recommended analgesic adjuvants (such as tranexemic acid, partial load walking and TENS) for pain control after THA surgery in the postoperative period.

While 49.2% of the participants did not prefer peripheral nerve block for pain control after THA surgery, the most preferred block among those who did was pericapsular nerve block (PENG) with a rate of 34.6% (**Figure 2**). Among the participants who were able to choose more than one option in their peripheral nerve block preferences, 27.6% were

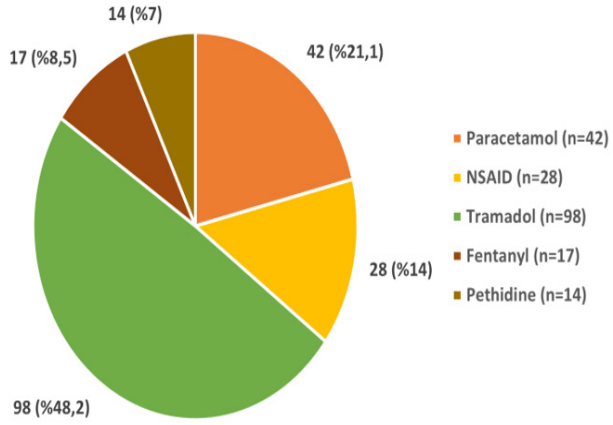


Figure 1. The most commonly administered intravenous analgesic drug in the postoperative period for pain control after total hip arthroplasty surgery
NSAID: Non-steroidal anti-inflammatory drug

Fascia iliaca block (FIB) and 16.5% were Femoral nerve block (FNB), Lateral femoral cutaneous nerve block (LFCSB) was performed in 13%, lumbar plexus block (LPB) in 4.5%, psoas compartment block (PCB) in 4%, and anterior quadratus lumborum block (AQLB) in 3%.

The rates of opioid use in the postoperative period are shown in **Figure 3**. While 93.5% of the participants preferred opioid

Table 4. General responses

n=199		**n (%)**
Do you recommend giving education about exercise and surgery in the preoperative period for pain control after total hip arthroplasty surgery?	Yes	148 (74.4%)
	No	51 (25.6%)
What is the most common method you use for pain control after total hip arthroplasty surgery?	General anesthesia	14 (7.1%)
	Spinal anesthesia	185 (92.9%)
Do you use intravenous paracetamol preoperatively and intraoperatively for pain control after total hip arthroplasty surgery?	Yes	146 (73.3%)
	No	53 (26.6%)
Do you use NSAIDs or COX2 selective inhibitors preoperatively and intraoperatively for pain control after total hip arthroplasty surgery?	Yes	129 (64.8%)
	No	69 (35.2%)
Do you use single dose intravenous dexamethasone preoperatively and intraoperatively for pain control after total hip arthroplasty surgery?	Yes	52 (26.1%)
	No	147 (73.8%)
Do you add an intrathecal long-acting opioid (morphine 0.1 mg) to spinal anesthesia for pain control after total hip arthroplasty surgery?	Yes	67 (66.3%)
	No	32 (33.7%)
Would you prefer patient-controlled epidural analgesia for pain control after total hip arthroplasty surgery?	Yes	145 (72.9%)
	No	54 (27.1%)
Do you use local infiltration analgesia for pain control after total hip arthroplasty surgery?	Yes	37 (18.6%)
	No	162 (81.4%)
Do you recommend analgesic adjuvants (such as tranexemic acid, partial load walking and TENS) in the postoperative period for pain control after total hip arthroplasty surgery?	Yes	14 (7.1%)
	No	185 (92.9%)
Do you follow PROSPECT recommendations?	Yes	86 (43.2%)
	No	113 (56.7%)

NSAID: Non-steroidal anti-inflammatory drug, COX2: Cyclooxygenase, TENS: Transcutaneous Electrical Nerve Stimulation, PROSPECT: Procedure-specific pain management. * n and % categorized as columns

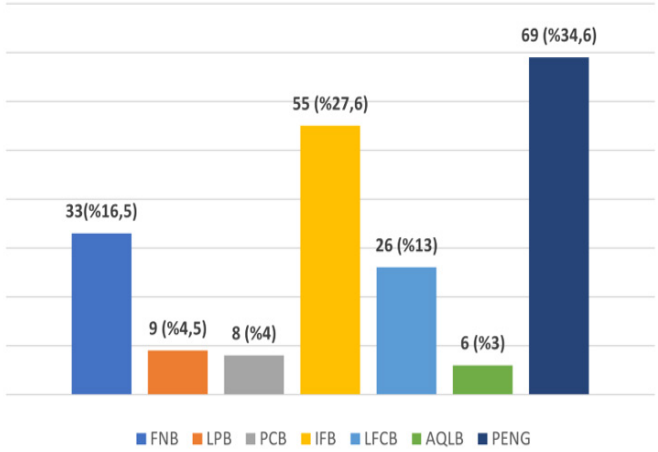


Figure 2. Peripheral nerve blocks preferred for pain control after total hip arthroplasty surgery
FNB: Femoral nerve block, LPB: Lumbar plexus block, PCB: Psoas compartment block, IFB: Iliac fascia block, LFCNB: Lateral femoral cutaneous nerve block, AQLB: Anterior quadratus lumborum block, PENG: Pericapsular nerve block

use in the postoperative period, it was observed that 35.7% of those who used opioids preferred opioids when they could not provide adequate analgesia with other analgesic methods. Of the participants, 24.6% reported using them in combination with paracetamol or NSAIDs, while 19.1% reported using them in combination with regional techniques.

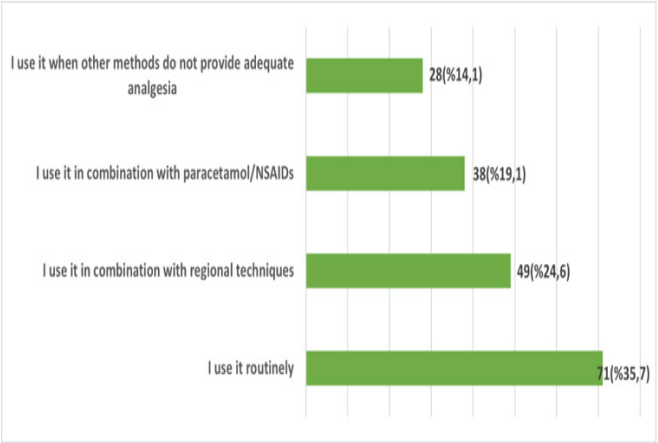


Figure 3. Postoperative opioid use for pain control after total hip arthroplasty surgery

Among the physicians who followed PROSPECT recommendations, 59.3% were research assistants, 19.1% were specialists, 6.5% were assistant professors, 9.5% were associate professors and 5.5% were professors (**Table 3**). Responses to PROSPECT recommendations according to professional title are summarized in **Table 5**. Among the physicians who preferred an intrathecal long-acting opioid (morphine 0.1 mg) in addition to spinal anesthesia, 56.7% were research assistants, 20.8% were specialists, 7.4% were assistant professors, 7.4% were associate professors and 7.4% were professors. Among the physicians who applied patient-controlled epidural analgesia after THA surgery, 58.6% were research assistants, 17% were specialists, 8.9% were assistant professors, 8.2% were associate professors and 6.8% were professors. When the preference rate of the most preferred analgesic drug tramadol in the postoperative period was compared according to the professional title, it was seen that it was preferred by research assistants with 63.9%.

When considering the duration of anesthesiology practice among physicians who follow PROSPECT recommendations, 46.2% had 0-10 years of experience, 27.6% had 10-20 years of experience, and 26.1% had over 20 years of experience (**Table 3**). Participants' responses to the survey questions based on their anesthetic practice are presented in **Table 6**. It was observed that 48.5% of physicians who preferred an additional long-acting opioid (morphine 0.1 mg) for spinal anesthesia had 0-10 years of anesthesia experience, 23.5% had 10-20 years, and 27.9% had 20 years or more. Among physicians who administered patient-controlled epidural analgesia after THA surgery, 44.8% had 0-10 years of anesthesia practice, 28.9% had 10-20 years, and 26.2% had 20 years or more. Among the 186 participants who used opioids in the postoperative period, a comparison based on the length of anesthetic practice showed that it was most preferred among physicians with 0-10 years of experience (46.2%).

When the type of hospital in which the physicians who followed PROSPECT recommendations after THA surgery were actively working was analyzed, it was seen that 24.1% were working in a university hospital, 16.1% in a training and research hospital, 39.7% in a public hospital and 20.1% in a private hospital (**Table 3**). The responses of the participants to PROSPECT recommendations according to the type of hospital where they were actively working are shown in **Table 7**. It was observed that 20.8% of the physicians who preferred an intrathecal long-acting opioid (morphine 0.1 mg) in addition to spinal anesthesia worked in a university hospital, 17.9% in a training and research hospital, 35.8% in a public hospital, and 25.3% in a private hospital. It was observed that 23.7% of the physicians who applied patient-controlled epidural analgesia after THA surgery worked in a university hospital, 16.7% in a training and research hospital, 41.2% in a state hospital, and 18.1% in a private hospital. Among the 186 participants who used opioids in the postoperative period, when compared according to the institution where they were actively working, it was seen that 38.7% preferred to use opioids in state hospitals. Among the analgesic drugs preferred in the postoperative period, tramadolu was the most preferred by the physicians working in state hospitals with a rate of 44.8%.

DISCUSSION

PROSPECT guideline of the European Society for Regional Anesthesia and Pain Management (ESRA) makes recommendations according to whether the intervention to be performed affects postoperative pain.[7]

A meta-analysis by Moyer et al.[8] found that preoperative exercise and education were beneficial in reducing postoperative pain and improving functional outcomes after THA surgery. PROSPECT recommends preoperative exercise and education (grade A). In our survey, 74.4% of the respondents recommended preoperative exercise and training.

Table 5. Responses according to title

		What is your professional title? (n (%))				
		Research Assistant	**Specialist Doctor**	**Assistant Professor**	**Associate Professor**	**Professor Doctor**
Do you recommend education about exercise and surgery in the preoperative period for pain control after total hip arthroplasty surgery?	Yes	83 (56.4%)	27 (18.3%)	11 (7.4%)	17 (11.5%)	9 (6.1%)
	No	35 (68.6%)	11 (21.5%)	2 (3.9%)	2 (3.9%)	1 (1.9%)
What is the most common method you use for pain control after total hip arthroplasty surgery?	General anesthesia	8 (57.1%)	3 (21.4%)	1 (7.1%)	2 (14.2%)	0 (0%)
	Spinal anesthesia	110 (59.4%)	35 (18.9%)	12 (6.4%)	17 (9.1%)	11 (5.9%)
Do you use intravenous paracetamol preoperatively or intraoperatively for pain control after total hip arthroplasty surgery?	Yes	84 (57.5%)	28 (19.1%)	11 (7.5%)	15 (10.2%)	8 (5.4%)
	No	34 (64.1%)	10 (18.8%)	2 (3.7%)	4 (7.5%)	3 (5.6%)
Do you use NSAIDs or COX2 selective inhibitors preoperatively or intraoperatively for pain control after total hip arthroplasty surgery?	Yes	74 (56.4%)	26 (19.8%)	10 (7.6%)	11 (8.3%)	10 (7.6%)
	No	44 (64.7%)	12 (17.6%)	3 (4.4%)	8 (11.7%)	1 (1.4%)
Do you use a single dose of intravenous dexamethasone before or during surgery for pain control after total hip arthroplasty?	Yes	28 (53.8%)	12 (23%)	1 (1.9%)	7 (13.4%)	4 (7.6%)
	No	90 (60.1%)	26 (17.6%)	12 (8.1%)	12 (8.1%)	7 (4.7%)
Do you add an intrathecal long-acting opioid (morphine 0.1 mg) to spinal anesthesia for pain control after total hip arthroplasty surgery?	Yes	38 (56.7%)	14 (20.8%)	5 (7.4%)	5 (7.4%)	5 (7.4%)
	No	80 (60.6%)	24 (18.1%)	8 (6%)	14 (10.6%)	6 (4.5%)
Would you prefer patient-controlled epidural analgesia for pain control after total hip arthroplasty surgery?	Yes	85 (58.6%)	25 (17.2%)	13 (8.9%)	12 (8.2%)	10 (6.8%)
	No	33 (61.1%)	13 (24%)	0 (0%)	7 (12.9%)	1 (1.8%)
Do you use local infiltration analgesia for pain control after total hip arthroplasty surgery?	Yes	21 (56.7%)	6 (16.2%)	4 (10.8%)	4 (10.8%)	2 (5.4%)
	No	97 (59.8%)	32 (19.7%)	9 (5.5%)	15 (9.2%)	9 (5.5%)
Do you recommend analgesic adjuvants (such as tranexemic acid, partial load walking and TENS) in the postoperative period for pain control after total hip arthroplasty surgery?	Yes	47 (57.3%)	19 (23.1%)	5 (6%)	6 (7.3%)	5 (6%)
	No	71 (60.6%)	19 (16.2%)	8 (6.8%)	13 (11.1%)	6 (5.1%)
Do you follow PROSPECT recommendations?	Yes	40 (47%)	16 (18.8%)	6 (7%)	14 (16.4%)	9 (10.5%)
	No	78 (69%)	22 (19.4%)	7 (6.1%)	4 (3.5%)	2 (1.7%)

NSAID: Non-steroidal anti-inflammatory drug, COX2: Cyclooxygenase, TENS: Transcutaneous Electrical Nerve Stimulation, PROSPECT: Procedure-specific pain management

Table 6. Responses based on professional experience

		How many years of anesthesiology practice do you have? (including residency) n (%)		
		0 to 10	**10 to 20**	**20 and above**
Do you recommend education about exercise and surgery in the preoperative period for pain control after total hip arthroplasty surgery?	Yes	68 (45.9%)	41 (27.7%)	39 (26.3%)
	No	25 (49%)	14 (27.4%)	12 (23.5%)
What is the most common method you use for pain control after total hip arthroplasty surgery?	General anesthesia	7 (43.7%)	3 (18.7%)	6 (37.5%)
	Spinal anesthesia	86 (46.9%)	52 (28.4%)	45 (24.5%)
Do you use intravenous paracetamol preoperatively or intraoperatively for pain control after total hip arthroplasty surgery?	Yes	68 (46.5%)	44 (30.1%)	34 (23.2%)
	No	25 (47.1%)	11 (20.7%)	17 (32%)
Do you use NSAIDs or COX2 selective inhibitors preoperatively or intraoperatively for pain control after total hip arthroplasty surgery?	Yes	57 (44.1%)	35 (27.1%)	37 (28.6%)
	No	36 (51.4%)	20 (28.5%)	14 (20%)
Do you use a single dose of intravenous dexamethasone before or during surgery for pain control after total hip arthroplasty?	Yes	27 (49%)	12 (21.8%)	16 (29%)
	No	66 (45.8%)	43 (29.8%)	35 (24.3%)
Do you add an intrathecal long-acting opioid (morphine 0.1 mg) for pain control after total hip arthroplasty surgery in patients under spinal anesthesia?	Yes	33 (48.5%)	16 (23.5%)	19 (27.9%)
	No	60 (45.8%)	39 (29.7%)	32 (24.4%)
Would you prefer patient-controlled epidural analgesia for pain control after total hip arthroplasty surgery?	Yes	65 (44.8%)	42 (28.9%)	38 (26.2%)
	No	28 (51.8%)	13 (24%)	13 (24%)
Do you use local infiltration analgesia for pain control after total hip arthroplasty surgery?	Yes	22 (59.4%)	6 (16.2%)	9 (24.3%)
	No	71 (43.8%)	49 (30.2%)	42 (25.9%)
Do you recommend analgesic adjuvants (such as tranexemic acid, partial load walking and TENS) in the postoperative period for pain control after total hip arthroplasty surgery?	Yes	46 (56%)	17 (20.7%)	19 (23.1%)
	No	47 (40.1%)	38 (32.4%)	32 (27.3%)
Do you follow PROSPECT recommendations?	Yes	44 (51.1%)	18 (20.9%)	24 (27.9%)
	No	49 (43.3%)	37 (32.7%)	27 (23.8%)

NSAID: Non-steroidal anti-inflammatory drug, COX2: Cyclooxygenase, TENS: Transcutaneous Electrical Nerve Stimulation, PROSPECT: Procedure-specific pain management

Table 7. Responses according to the type of hospital worked in

		What is the type of hospital you are actively working in? (n (%))			
		University Hospital	Ministry of Health Training and Research Hospital	State Hospital	Private Hospital
Do you recommend education about exercise and surgery in the preoperative period for pain control after total hip arthroplasty surgery?	Yes	34 (22.8%)	21 (14%)	60 (40.2%)	34 (22.8%)
	No	14 (28%)	11 (22%)	19 (38%)	6 (12%)
What is the most common method you use for pain control after total hip arthroplasty surgery?	General anesthesia	2 (14.2%)	5 (35.7%)	2 (14.2%)	5 (35.7%)
	Spinal anesthesia	46 (24.8%)	27 (14.5%)	77 (41.6%)	35 (18.9%)
Do you use intravenous paracetamol preoperatively or intraoperatively for pain control after total hip arthroplasty surgery?	Yes	32 (21.6%)	26 (17.5%)	56 (37.8%)	34 (22.9%)
	No	16 (31.3%)	6 (11.7%)	23 (45%)	6 (11.7%)
Do you use NSAIDs or COX2 selective inhibitors preoperatively or intraoperatively for pain control after total hip arthroplasty surgery?	Yes	28 (21.7%)	25 (19.3%)	49 (37.9%)	27 (20.9%)
	No	20 (28%)	7 (10%)	30 (42.8%)	13 (18.5%)
Do you use a single dose of intravenous dexamethasone before or during surgery for pain control after total hip arthroplasty?	Yes	10 (19.2%)	11 (21.1%)	18 (34.6%)	13 (25%)
	No	38 (25.8%)	21 (14.2%)	61 (41.4%)	27 (18.3%)
Do you add an intrathecal long-acting opioid (morphine 0.1 mg) for pain control after total hip arthroplasty surgery in patients under spinal anesthesia?	Yes	14 (20.8%)	12 (17.9%)	24 (35.8%)	17 (25.3%)
	No	34 (25.7%)	20 (15.1%)	55 (41.6%)	23 (17.4%)
Would you prefer patient-controlled epidural analgesia for pain control after total hip arthroplasty surgery?	Yes	34 (23.7%)	24 (16.7%)	59 (41.2%)	26 (18.1%)
	No	14 (25%)	8 (14%)	20 (35.7%)	14 (25%)
Do you use local infiltration analgesia for pain control after total hip arthroplasty surgery?	Yes	8 (21.6%)	11 (29.7%)	12 (32.4%)	6 (16.2%)
	No	40 (24.6%)	21 (12.9%)	67 (41.3%)	34 (20.9%)
Do you recommend analgesic adjuvants (such as tranexemic acid, partial load walking and TENS) in the postoperative period for pain control after total hip arthroplasty surgery?	Yes	13 (15.8%)	15 (18.2%)	34 (41.4%)	20 (24.3%)
	No	35 (29.9%)	17 (14.5%)	45 (38.4%)	20 (17%)
Do you follow PROSPECT recommendations?	Yes	11 (12.7%)	12 (13.9%)	36 (41.8%)	27 (31.3%)
	No	37 (32.7%)	20 (17.6%)	43 (38%)	13 (11.5%)

NSAID: Non-steroidal anti-inflammatory drug, COX2: Cyclooxygenase, TENS: Transcutaneous Electrical Nerve Stimulation, PROSPECT: Procedure-specific pain management

When we look at our survey results, it is seen that spinal anesthesia is the preferred anesthesia method for pain control after THA surgery with a rate of 92.9%. We think that the reason for this high preference is to avoid the side effects of general anesthesia such as nausea-vomiting and agitation rather than postoperative pain control. Although spinal anesthesia has been reported to favorably affect other postoperative outcomes compared with general anesthesia, there is insufficient evidence to support one anesthesia technique over the other in terms of postoperative analgesic benefits.[9] Therefore, both are among the PROSPECT recommendations for pain control after THA surgery (grade A).

In the literature, there are studies reporting that the use of iv paracetamol reduces opioid consumption after THA surgery.[10,11] Westrich et al.[12] found no difference in postoperative pain outcomes in their study comparing iv and oral paracetamol administration. In two separate studies comparing the combination of paracetamol with both ibuprofen[13] and parecoxib[14] with the use of these drugs alone, it was reported that the combined use did not result in clinically significant improvement. According to these studies, paracetamol has a limited effect when used in combination with COX-2 selective inhibitors or NSAIDs, but is generally recommended as part of basic postoperative analgesia due to its minor side effects.

There are studies showing that NSAID administration reduces postoperative pain. Gombotz et al.[15] reported that regular postoperative iv infusions of diclofenac and orphenadrine reduced postoperative opioid patient-controlled analgesia (PCA) consumption. McQuay et al.[16] showed that the combination of 25 mg oral dexketoprofen and 75 mg tramadol was superior for postoperative pain control in a study comparing the use of both drugs alone.

In one of the two studies on the effects of preoperative COX-2 selective inhibitors in reducing postoperative pain, oral etoricoxib[17] was administered 2 hours before surgery and iv parecoxib or oral celecoxib[18] was administered 1 hour before surgery. It was found to be associated with significantly lower postoperative pain scores and morphine consumption in those who received COX-2 selective inhibitors compared to placebo.

The PROSPECT guideline recommends the use of paracetamol, NSAIDs and COX-2 selective inhibitors in all perioperative periods for pain control after THA surgery (grade A). In our survey, the majority of participants used paracetamol, NSAIDs and COX-2 selective inhibitors preoperatively and intraoperatively.

There are many studies showing the benefit of glucocorticoids among non-opioid adjunctive analgesics for postoperative pain control. Perioperative 125 mg methylprednisolone reduced 24-hour pain scores compared with placebo.[19] Backes et al.[20] showed that perioperative 10 mg dexamethasone reduced mean visual pain scores (VAS), reduced opioid consumption in the first 24 hours, provided early mobilization and shorter

hospital stay. For pain control after THA surgery, a single dose of iv dexamethasone preoperatively or intraoperatively is recommended by PROSPECT (grade A). Other non-opioid adjunctive analgesics such as pregabalin and ketamine are not recommended due to lack of sufficient evidence. In our survey, the rate of single dose iv dexamethasone use was low at 26.1%. We think that this is due to the insufficient rate of follow-up of PROSPECT recommendations.

There was no consensus among PROSPECT members regarding the use of intrathecal morphine as an adjunct to spinal anesthesia (grade D). This is because some studies[21] reported analgesia for at least 24 hours postoperatively and limited side effects at small doses (≤ 0.1 mg morphine), while others[22] reported pruritus and postoperative nausea and vomiting, which may delay walking and oral intake and affect patient satisfaction.

According to PROSPECT, epidural analgesia is effective but not recommended for lower extremity surgery due to well-known side effects such as limb weakness, bladder dysfunction and delayed mobilization.[23] Our survey results showed that the majority of participants (72.9%) preferred patient-controlled epidural analgesia for pain control after THA surgery.

The International Consensus on Anaesthesia-Related Outcomes after Surgery (ICAROS) group recommends peripheral nerve blocks in THA to reduce respiratory failure, cognitive dysfunction, cardiac complications, surgical site infections, blood transfusions, thromboembolism, and intensive care unit admissions, and they also note that nerve blocks show stronger effects when combined with general anesthesia compared to neuraxial anesthesia.[24] However, new fascial plane blocks such as the PENG block were not included in the analysis. In the PROSPECT guideline, only certain blocks are mentioned and not recommended due to limited procedure-specific evidence or side effects such as delayed mobilization and motor block. The majority of the respondents in our survey practiced the PENG block, which is the most recent block among peripheral blocks. More studies are needed to include peripheral blocks in PROSPECT recommendations.

In a meta-analysis comparing LIA with epidural, Yan et al.[25] found no significant difference between the groups in terms of pain with movement 48-72 hours after surgery, but found less pain at 24 hours in the epidural group. In a large meta-analysis[26] including 35 randomized controlled trials and 2296 patients, LIA, peripheral nerve block and placebo were compared. The LIA group had lower postoperative pain scores and opioid consumption at 24 hours postoperatively compared to placebo. Single injection LIA, previously not recommended, is now included in the latest PROSPECT recommendations based on supporting studies and is reported to have analgesic effect with no side effects (grade D). The results of our survey showed that the use of LIAs for pain control after THA surgery was low (18.6%) among respondents. Again, we think that this is due to the fact that PROSPECT recommendations are not being followed sufficiently.

PROSPECT recommendations for postoperative interventions include only the use of rescue opioids (grade D). Other postoperative interventions such as tranexemic acid, partial load walking and the use of TENS are not recommended by PROSPECT due to insufficient evidence. Musclow et al.[27] compared a PCA regimen of paracetamol/NSAIDs/morphine to placebo by adding 30 mg of oral modified-release morphine every 12 hours. Modified-release morphine was not proven to be effective on pain scores and was even associated with significantly more opioid-related side effects. Among the physicians who participated in our survey, routine opioid use in the postoperative period for pain control after THA surgery was 14%, and 35.6% of opioid users used opioids when other analgesic methods failed to provide adequate analgesia. Although postoperative opioid use by PROSPECT is in the low recommendation group, it is seen that tramadol is the most commonly used agent used by the participants in our survey.

In the most recent literature update in 2021, some recommendations for pain management after THA surgery were changed.[7] For example, previously recommended approaches such as FNB, LPB and epidural analgesia are no longer recommended due to the availability of evidence supporting better and safer alternatives such as FIB and LIA. LIA and dexamethasone were previously not recommended due to inconsistent evidence, but in the latest update, single injection LIA and 8-10 mg iv dexamethasone are now recommended based on supporting studies.

In our survey, 43.2% of the participants followed PROSPECT recommendations for pain management after THA surgery. While this situation is not very different when the hospitals worked in are taken into account, it is seen that the rate of following the recommendations increases as the academic career increases among the respondents.

Limitations

The limitation of this study is that the questionnaire was sent via e-mail only to Anesthesiology and Reanimation physicians who are members of TARD. If there is a chance to reach non-member physicians, a survey with a wider participation can be created.

CONCLUSION

Since THA surgery is mostly performed in the elderly population with comorbid diseases, there is no consensus among Anesthesiology and Reanimation physicians about postoperative analgesia methods. There has been a recent increase in peripheral block applications in these patients due to anticoagulant use. Despite their side effects, opioids are still widely preferred. We think that postoperative pain management trainings in THA surgery should be increased and PROSPECT recommendations should be disseminated in these trainings.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study was approved by the Giresun Training and Research Hospital Ethics Committee (Date: 19.03.2025, Decision No: 19.03.2025/05).

Informed Consent

All patients signed and free and informed consent form.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Lee B, Schug SA, Joshi GP, Kehlet H; PROSPECT Working Group. Procedure-specific pain management (PROSPECT)-an update. *Best Pract Res Clin Anaesthesiol*. 2018;32(2):101-111. doi:10.1016/j.bpa.2018.06.012
2. Lavand'homme PM, Kehlet H, Rawal N, Joshi GP; PROSPECT Working Group of the European Society of Regional Anaesthesia and Pain Therapy (ESRA). Pain management after total knee arthroplasty: procedure specific postoperative pain management recommendations. *Eur J Anaesthesiol*. 2022;39(9):743-757. doi:10.1097/EJA.0000000000001691
3. Højer Karlsen AP, Geisler A, Petersen PL, Mathiesen O, Dahl JB. Postoperative pain treatment after total hip arthroplasty: a systematic review. *Pain*. 2015;156(1):8-30. doi:10.1016/j.pain.0000000000000003
4. Joshi GP, Kehlet H. Postoperative pain management in the era of ERAS: an overview. *Best Pract Res Clin Anaesthesiol*. 2019;33(3):259-267. doi:10.1016/j.bpa.2019.07.016
5. Safa B, Trinh H, Lansdown A, et al. Ultrasound-guided suprainguinal fascia iliaca compartment block and early postoperative analgesia after total hip arthroplasty: a randomized controlled trial. *Br J Anaesth*. 2024;133(1):146-151. doi:10.1016/j.bja.2024.04.019
6. Geisler A, Dahl JB, Thybo KH, et al. Pain management after total hip arthroplasty at five different Danish hospitals: a prospective, observational cohort study of 501 patients. *Acta Anaesthesiol Scand*. 2019;63(7):923-930. doi:10.1111/aas.13349
7. Anger M, Valovska T, Beloeil H, et al. PROSPECT guideline for total hip arthroplasty: a systematic review and procedure-specific postoperative pain management recommendations. *Anaesthesia*. 2021;76(8):1082-1097. doi:10.1111/anae.15498
8. Moyer R, Ikert K, Long K, Marsh J. The value of preoperative exercise and education for patients undergoing total hip and knee arthroplasty: a systematic review and meta-analysis. *JBJS Rev.* 2017;5(12):e2. doi:10.2106/JBJS.RVW.17.00015
9. Memtsoudis SG, Cozowicz C, Bekeris J, et al. Anaesthetic care of patients undergoing primary hip and knee arthroplasty: consensus recommendations from the International Consensus on Anaesthesia-Related Outcomes after Surgery group (ICAROS) based on a systematic review and meta-analysis. *Br J Anaesth*. 2019;123(3):269-287. doi:10.1016/j.bja.2019.05.042
10. Sinatra RS, Jahr JS, Reynolds L, et al. Intravenous acetaminophen for pain after major orthopedic surgery: an expanded analysis. *Pain Pract*. 2012;12(5):357-365. doi:10.1111/j.1533-2500.2011.00514.x
11. Takeda Y, Fukunishi S, Nishio S, Yoshiya S, Hashimoto K, Simura Y. Evaluating the effect of intravenous acetaminophen in multimodal analgesia after total hip arthroplasty: a randomized controlled trial. *J Arthroplasty*. 2019;34(6):1155-1161. doi:10.1016/j.arth.2019.02.033
12. Westrich GH, Birch GA, Muskat AR, et al. Intravenous vs oral acetaminophen as a component of multimodal analgesia after total hip arthroplasty: a randomized, blinded trial. *J Arthroplasty*. 2019;34(7S):S215-S220. doi:10.1016/j.arth.2019.02.030
13. Thybo KH, Hägi-Pedersen D, Dahl JB, et al. Effect of combination of combination of paracetamol (acetaminophen) and ibuprofen vs either alone on patient-controlled morphine consumption in the first 24 hours after total hip arthroplasty: the PANSAID randomized clinical trial. *JAMA*. 2019;321(6):562-571. doi:10.1001/jama.2018.22039
14. Camu F, Borgeat A, Heylen RJ, Viel EJ, Boye ME, Cheung RY. Parecoxib, propacetamol, and their combination for analgesia after total hip arthroplasty: a randomized non-inferiority trial. *Acta Anaesthesiol Scand*. 2017;61(1):99-110. doi:10.1111/aas.12841
15. Gombotz H, Lochner R, Sigl R, Blasl J, Herzer G, Trimmel H. Opiate sparing effect of fixed combination of diclophenac and orphenadrine after unilateral total hip arthroplasty: a double-blind, randomized, placebo-controlled, multi-centre clinical trial. *Wien Med Wochenschr*. 2010;160(19-20):526-534. doi:10.1007/s10354-010-0829-7
16. McQuay HJ, Moore RA, Berta A, et al. Randomized clinical trial of dexketoprofen/tramadol 25 mg/75 mg in moderate-to-severe pain after total hip arthroplasty. *Br J Anaesth*. 2016;116(2):269-276. doi:10.1093/bja/aev457
17. Renner B, Walter G, Strauss J, Fromm MF, Zacher J, Brune K. Preoperative administration of etoricoxib in patients undergoing hip replacement causes inhibition of inflammatory mediators and pain relief. *Eur J Pain*. 2012;16(6):838-848. doi:10.1002/j.1532-2149.2011.00062.x
18. Ittichaikulthol W, Prachanpanich N, Kositchaiwat C, Intapan T. The post-operative analgesic efficacy of celecoxib compared with placebo and parecoxib after total hip or knee arthroplasty. *J Med Assoc Thai*. 2010;93(8):937-942.
19. Lunn TH, Andersen LØ, Kristensen BB, et al. Effect of high-dose preoperative methylprednisolone on recovery after total hip arthroplasty: a randomized, double-blind, placebo-controlled trial. *Br J Anaesth*. 2013;110(1):66-73. doi:10.1093/bja/aes345
20. Backes JR, Bentley JC, Politi JR, Chambers BT. Dexamethasone reduces length of hospitalization and improves postoperative pain and nausea after total joint arthroplasty: a prospective, randomized controlled trial. *J Arthroplasty*. 2013;28(8 Suppl):11-17. doi:10.1016/j.arth.2013.05.041
21. Albrecht E, Bayon V, Hirotsu C, Ja'bari A, Heinzer R. Intrathecal morphine and sleep apnoea severity in patients undergoing hip arthroplasty: a randomized, controlled, triple-blinded trial. *Br J Anaesth*. 2020;125(5):811-817. doi:10.1016/j.bja.2020.07.052
22. Kuchálik J, Granath B, Ljunggren A, Magnuson A, Lundin A, Gupta A. Postoperative pain relief after total hip arthroplasty: a randomized, double-blind comparison between intrathecal morphine and local infiltration analgesia. *Br J Anaesth*. 2013;111(5):793-799. doi:10.1093/bja/aet248
23. Spreng UJ, Dahl V, Hjall A, Fagerland MW, Ræder J. High-volume local infiltration analgesia combined with intravenous or local ketorolac+morphine compared with epidural analgesia after total knee arthroplasty. *Br J Anaesth*. 2010;105(5):675-682. doi:10.1093/bja/aeq232
24. Memtsoudis SG, Cozowicz C, Bekeris J, et al. Peripheral nerve block anesthesia/analgesia for patients undergoing primary hip and knee arthroplasty: recommendations from the International Consensus on Anesthesia-Related Outcomes after Surgery (ICAROS) group based on a systematic review and meta-analysis of current literature. *Reg Anesth Pain Med*. 2021;46(11):971-985. doi:10.1136/rapm-2021-102750
25. Yan H, Cang J, Xue Z, Lu J, Wang H. Comparison of local infiltration and epidural analgesia for postoperative pain control in total knee arthroplasty and total hip arthroplasty: a systematic review and meta-analysis. *Bosn J Basic Med Sci*. 2016;16(4):239-246. doi:10.17305/bjbms.2016.1072
26. Jiménez-Almonte JH, Wyles CC, Wyles SP, et al. Is local infiltration analgesia superior to peripheral nerve blockade for pain management after total hip arthroplasty: a network meta-analysis. *Clin Orthop Relat Res*. 2016;474(2):495-516. doi:10.1007/s11999-015-4619-9
27. Musclow SL, Bowers T, Vo H, Glube M, Nguyen T. Long-acting morphine following hip or knee replacement: a randomized, double-blind, placebo-controlled trial. *Pain Res Manag*. 2012;17(2):83-88. doi:10.1155/2012/704932

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1751000

Anatolian Curr Med J. 2025;7(5):687-694

Effectiveness of artificial intelligence algorithms in predicting progression-free survival in epithelial ovarian cancer patients

Aysun Alcı[1], Necim Yalçın[2], Mustafa Gökkaya[3], Gülsüm Ekin Sarı[4], Fatih İkiz[5], Işın Üreyen[4], Tayfun Toptaş[4]

[1]*Department of Gynecologic Oncology, Kahramanmaraş Necip Fazıl City Hospital, Kahramanmaraş, Turkiye*

[2]*Department of Gynecologic Oncology, Batman Training and Research Hospital, Batman, Turkiye*

[3]*Department of Gynecologic Oncology, Gaziantep City Hospital, Gaziantep, Turkiye*

[4]*Department of Gynecologic Oncology, Antalya Training and Research Hospital, University of Health Sciences, Antalya, Turkiye*

[5]*Department of Emergency Medicine, Beyhekim Training and Research Hospital, University of Health Sciences, Konya, Turkiye*

Cite this article as: Alcı A, Yalçın N Gökkaya M, et al. Effectiveness of artificial intelligence algorithms in predicting progression-free survival in epithelial ovarian cancer patients. *Anatolian Curr Med J.* 2025;7(5):687-694.

Received: 25.07.2025 • **Accepted**: 08.09.2025 • **Published**: 15.09.2025

ABSTRACT

Aims: This study aimed to assess the predictive performance of artificial intelligence–based models in estimating progression-free survival (PFS) in patients with epithelial ovarian cancer and to compare various interpretable machine learning approaches.

Methods: Between January 2015 and December 2020, a total of 167 patients who underwent surgical intervention at the Gynaecological Oncology Department of Antalya Training and Research Hospital were retrospectively included in the study if their data were complete. Clinical data were analysed, and the dataset was randomly divided into a training group (n=117; 75%) and a validation group (n=42; 25%). A machine learning (ML) analysis was conducted using the eight most relevant and widely applied algorithmic models for this study design. Model development time, mean absolute error (MAE), root mean square error (RMSE), and correlation coefficient (CC) were evaluated.

Results: Random Forest demonstrated the highest accuracy (MAE=16.45, CC=0.571, RMSE=20.98, time=0.03) and thus became the focus of subsequent analyses. Other algorithms included Linear Regression, Bootstrap Aggregating, Additive Regression, Random Committee, and Regression by Discretization (CC=0.533, 0.492, 0.449, 0.408, and 0.382, respectively). For Random Forest, a moderate correlation was observed between actual and predicted PFS values (CC=0.4–0.6), indicating moderate predictive performance.

Conclusion: The findings of this study demonstrate that machine learning models, particularly Random Forest, can achieve moderate yet clinically relevant prognostic performance based on routinely collected clinical data. In particular, Random Forest demonstrates potential clinical value in guiding patient follow-up strategies and supporting individualized management in ovarian cancer, although further research is required to enhance its clinical validity and applicability.

Keywords: Artificial intelligence, deep learning in gynecologic oncology, epithelial ovarian neoplasms

INTRODUCTION

Ovarian cancer is the eighth most common cancer among women worldwide, accounting for approximately 3.7% of diagnoses and 4.7% of cancer-related deaths, although its incidence varies significantly across regions.[1] The current standard treatment for epithelial ovarian cancer consists of maximal cytoreductive surgery followed by platinum-based chemotherapy, with the possible addition of maintenance therapies such as bevacizumab and/or Poly (ADP-ribose) polymerase (PARP-1) inhibitors.[2,3] Conventional prognostic tools rely on parameters such as tumor stage, histology, patient age, comorbidities, and the extent of cytoreduction; however, these factors often fail to capture the complexity and heterogeneity of the disease.[4] Recent advances in machine learning (ML) have introduced novel approaches for improving prognostic accuracy in oncology. By applying sophisticated analytical methods to large and complex datasets, ML can identify patterns that remain undetected by traditional statistical techniques. The effectiveness of methods such as supervised learning, neural networks, and ensemble approaches has already been demonstrated in various cancer types, highlighting their potential to outperform conventional models in outcome prediction. Consequently, the integration of artificial intelligence (AI) and ML into oncology holds promise for enhancing diagnostic and prognostic accuracy and for enabling more personalized and effective treatment strategies.[5-7]

Corresponding Author: Aysun Alcı, aysun_alci@hotmail.com

With the increasing emphasis on personalized medicine, the need for reliable predictive tools tailored to individual patient characteristics has become more pressing. This is particularly important for heterogeneous diseases such as epithelial ovarian cancer, where traditional statistical models are still difficult to apply in practice. In this context, the aim of our study was to assess the potential of AI based models to provide patient-specific prognostic information, which is critical for the clinical management of these tumors.[8,9] Specifically, we compared the performance of ML methods in predicting progression-free survival (PFS) in patients with epithelial ovarian cancer using preoperative, intraoperative, and postoperative clinical variables. This evaluation seeks to provide a more comprehensive understanding of both the advantages and limitations of AI driven approaches in clinical oncology, and to offer insights that may inform future strategies for clinical practice.

METHODS

Ethics

This retrospective analysis was approved by the Scientific Ethics Committee for Medical Researches at Antalya Training and Research Hospital in Turkiye (Date: 07.11.2024, Decision No: 17/2), was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki.

Study Design and Patients

The data utilized in this study were fully anonymized prior to analysis and contained no personally identifiable information. We reviewed the records of consecutive adult patients (aged 18 years and above) who underwent surgery for epithelial ovarian cancer at our institution between January 2015 and December 2020. Patients with other gynaecological malignancies or with benign histopathological findings were excluded from the analysis. Individuals diagnosed with nonovarian gynaecological malignancies or benign pathological entities were not included in the final analysis.

Data Collection

Demographic and clinical characteristics were recorded for all eligible patients, including age, diabetes mellitus status, Eastern Cooperative Oncology Group (ECOG) performance score, and the presence of significant cardiovascular or pulmonary comorbidities. The extent of intra-abdominal disease was documented in accordance with standard clinical guidelines, and any evidence of extra-abdominal spread on preoperative imaging was noted. In line with evidence-based recommendations from randomized controlled trials,[10,11] patients considered suitable for interval debulking surgery initially received neoadjuvant chemotherapy (NACT). Surgical cytoreduction was performed approximately 21 days after the final NACT cycle. The total number of NACT cycles, typically three, four, or six, was carefully recorded for each patient.

Detailed intraoperative findings were also documented, including the type and extent of surgical procedures performed (e.g., multiorgan resections when applicable), the need for intraoperative blood transfusion, and the degree of cytoreduction achieved, categorized according to residual tumor size after debulking. Postoperative outcomes were assessed by recording intensive care unit (ICU) admissions and length of hospital stay.

Preoperative serum CA-125 levels were obtained for all patients. Pathological and adjuvant treatment data were likewise collected, including International Federation of Gynecology and Obstetrics (FIGO) stage, histological subtype, and the number of lymph nodes resected. Receipt of adjuvant chemotherapy and the number of cycles administered were also recorded. Disease status at the first post-treatment evaluation (e.g., no evidence of disease vs. residual disease) was assessed. Postoperative complications were graded using the Clavien–Dindo classification system.[12] Finally, the interval between debulking surgery and initiation of adjuvant therapy was documented for each patient.

Machine Learning Model Development

The dataset was randomly divided into two subsets: approximately 75% of the patient records were allocated for model training, and the remaining 25% were reserved for testing. To ensure an optimal partitioning strategy, multiple train-test ratios were evaluated (10%, 20%, 25%, 40%, and 50%). Among these, the 25% test set provided the best balance between model development and evaluation, resulting in 117 patients in the training set and 42 in the test set. Model construction including feature selection and algorithm training was performed exclusively on the training dataset, while the test set was retained for independent validation. The distribution of outcome classes (group 1 and group 2) was assessed using z-tests, which confirmed no statistically significant imbalance between subsets. To ensure the stability and generalizability of our models, we evaluated them using a repeated random sub-sampling validation strategy. We performed 100 iterations of partitioning the dataset. In each iteration, the data was randomly split into a training set (75% of patients, n=117) and a test set (25% of patients, n=42). To prevent distributional bias, the splits were stratified to maintain the same proportion of outcome classes (group 1 and group 2) in both the training and test sets as in the original cohort. Missing values were handled internally by the classifier's default method, which distributes instances with unknown values fractionally across the branches of the decision trees based on the observed training data distribution. The feature of importance have defined with Shapley Additive Explanations (SHAP) values, using Python version 3.14 (**Figure 1, 2**).

Eight ML algorithms were applied, selected based on their prevalence in the literature and relevance to the classification task, and implemented using the Waikato Environment for Knowledge Analysis (WEKA), version 3.8.6. Following training, predictive performance was evaluated on the test set using classification accuracy along with additional indicators of predictive strength. To identify the most effective model, performance metrics including mean absolute error (MAE), root mean squared error (RMSE), and Pearson's correlation coefficient (CC) were calculated, while model calibration and fit were examined through statistical comparisons between

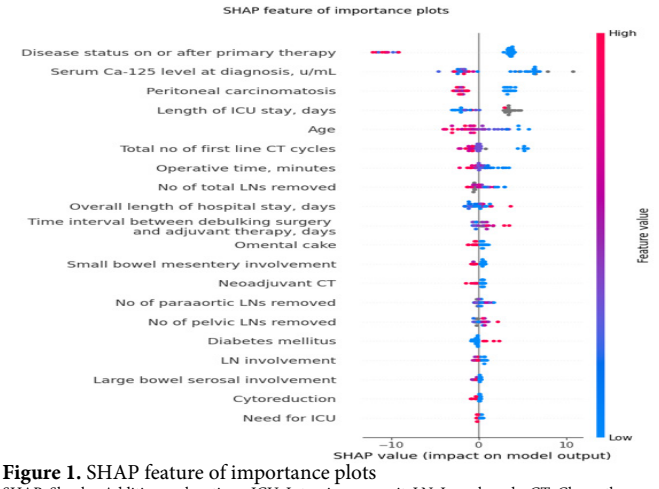


Figure 1. SHAP feature of importance plots
SHAP: Shapley Additive explanations, ICU: Intensive care unit, LN: Lymph node, CT: Chemotherapy

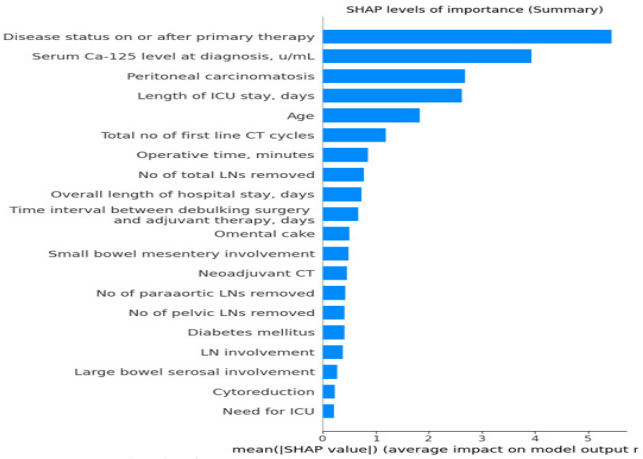


Figure 2. SHAP levels of importance
SHAP: Shapley Additive explanations, ICU: Intensive care unit, LN: Lymph node, CT: Chemotherapy

predicted and observed results in the test cohort. Random Forest, the best-performing algorithm, was carried out with 100 trees and a tree depth value of 10.

Statistical Analysis

All supplementary statistical analyses were performed using IBM SPSS Statistics (version 27.0; IBM Corp., Chicago, IL, USA), underscoring the use of this software for rigorous data analysis. The normality of continuous data distributions was evaluated with the Kolmogorov–Smirnov test, supplemented by visual inspection methods such as histograms, Q–Q plots, and examination of skewness and kurtosis. Categorical variables were presented as frequencies with corresponding percentages for clarity. Pearson's CC was employed to examine associations between continuous variables. To evaluate the concordance between predicted and observed values of the primary outcome, a paired-samples t-test was performed. All ML computations were executed on a Windows 11 system equipped with an Intel Core i7 CPU, 16 GB RAM, and an NVIDIA GeForce GTX 1660 Ti graphics card (8 GB of memory). A two-tailed significance level of 5% (α=0.05) was applied to all statistical tests, and p-values below this threshold were considered statistically significant.

RESULTS

Figure 3 shows the inclusion of 167 patients who underwent surgery for epithelial ovarian carcinoma during the study period. As presented in Table 1, 2, the mean patient age was 58±11 years. Notably, 36.5% of the cohort (n=61) had stage III disease, while 74.9% (n=125) exhibited high-grade tumor histology. A total of eight widely used ML algorithms were tested, and their predictive performance was evaluated. The algorithms utilised in this study were Random Forest, Multilayer Perceptron, Linear Regression, Support Vector Regression, Additive Regression, Bootstrap Aggregating (bagging), Random Committee and Regression by Discretisation. The Multilayer Perceptron and Support Vector Regression algorithms exhibited the lowest performance in terms of CC (0.1543, 0.1997). Following a rigorous evaluation process, the Random Forest algorithm was identified as the most effective algorithm and thus became the focus of subsequent research. The CC of the Random Forest algorithm was 0.5731, with a MAE of 16.45 and a RMSE of 20.98. The time required to create the model was 0.03 seconds. The remaining algorithms were Linear Regression, Bootstrap Aggregating (bagging), Additive Regression, Random Committee, and Regression by Discretization (CCs: 0.5326, 0.4915, 0.4491, 0.4077, 0.3817) (Table 3). A statistical analysis of actual and predicted PFS was performed to determine the success rate of the best performing Random Forest algorithm. A moderately significant correlation was found between actual and predicted PFS ($p<0.001$ and CC=0.573). In addition, an analysis of the difference between the actual and predicted PFS values was performed and no statistically significant difference was found (the difference between the actual and predicted values was very small and the p-value was greater than 0.946) (Table 4, Figure 3). This study highlights that the Random Forest algorithm provides the highest prediction accuracy compared to the other models tested. The moderate correlation observed between actual and PFS values (CC=0.573, $p<0.001$), despite the model demonstrating meaningful predictive capacity, indicates that its performance is not yet optimal for clinical application. Importantly, the absence of statistically significant differences between actual and predicted values (p=0.946) further supports the model's validity. These findings suggest that Random Forest could serve as a promising foundation for clinical prognostic tools in epithelial ovarian cancer. However, future studies with larger cohorts, more diverse datasets, and refined parameter optimisation are necessary to improve prediction accuracy and clinical applicability.

The Bland–Altman analysis demonstrated an overall acceptable agreement between the actual and predicted PFS values. The mean difference was close to zero, indicating the absence of a systematic bias in the predictions. Most of the data points were within the 95% limits of agreement, reflecting a generally reliable concordance between the model outputs and observed outcomes. However, a wider spread of differences was noted at higher mean PFS values, suggesting reduced reliability of the model in patients with longer

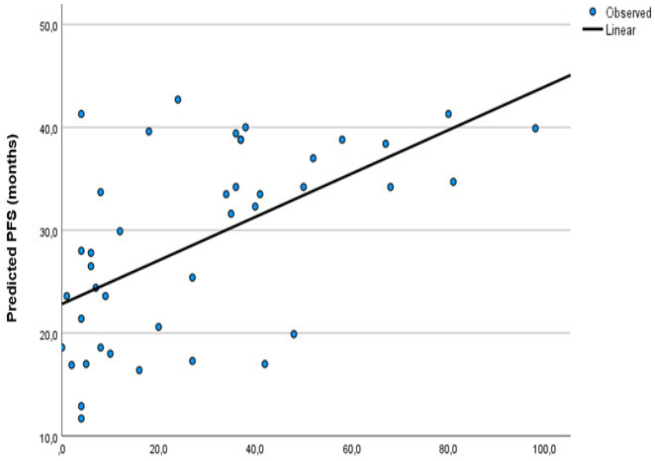


Figure 3. Correlation relationship between the actual and predicted PFS values
PFS: Progression-free survival

survival times. These findings imply that while the Random Forest algorithm provides statistically consistent predictions, its clinical applicability may be limited, particularly for cases with extended PFS (**Figure 4**).

DISCUSSION

To the best of our knowledge, this study is among the few that comprehensively incorporate demographic data, intraoperative and perioperative findings, and adjuvant treatment responses to evaluate the predictive power of ML models for PFS in patients undergoing surgery for epithelial ovarian cancer. Among the algorithms tested, Random Forest demonstrated the highest performance, while Multilayer Perceptron and Support Vector Regression showed the lowest performance, with CCs of 0.1543 and 0.1997, respectively. Random Forest achieved the best results, with a CC of 0.5731,

Table 1. Summary of the distribution of qualitative parameters

Parameter		n (%)	Parameter		n (%)
ECOG-PS	0-1	129 (77.25%)	Large bowel serosal invasion	No	97 (58.08%)
	≥2	38 (22.75%)		Localized foci	40 (23.95%)
Major cardiac comorbidities	No	128 (76.65%)		Diffuse. military	30 (17.96%)
	Yes	39 (23.35%)	Large bowel mesentery invasion	No	75 (44.91%)
Major pulmonary comorbidities	No	153 (91.62%)		Localized foci	54 (32.34%)
	Yes	14 (8.38%)		Diffuse. military	38 (22.75%)
Diabetes mellitus	No	138 (82.63%)	Spleen metastasis	No	149 (89.22%)
	Yes	29 (17.37%)		Yes	18 (10.78%)
Neoadjuvant KT	No	109 (65.27%)	Liver metastasis	No	150 (89.82%)
	3 cycles	22 (13.17%)		Any surface lesion	11 (6.59%)
	4 cycles	25 (14.97%)		Parencyhmal	6 (3.59%)
	≥6 cycles	11 (6.59%)	Pleural effusion	No	133 (79.64%)
Ascite	No	108 (64.67%)		Yes	34 (20.36%)
	Small volume	27 (16.17%)	Pleural or pulmonary nodule	No	157 (94.01%)
	Large volume	32 (19.16%)		Yes	10 (5.99%)
Omental cake	No	106 (63.47%)	Mediastinal and or paracardiac LN	No	145 (86.83%)
	Yes	61 (36.53%)		Yes	22 (13.17%)
Peritoneal carcinomatosis	No	67 (40.12%)	Inguinal LN	No	164 (98.2%)
	Localized foci	34 (20.36%)		Yes	3 (1.8%)
	Diffuse, military	66 (39.52%)	Supraklavikular LN	No	164 (98.2%)
Diyaphragmatic disease	No	120 (71.86%)		Yes	3 (1.8%)
	Localized foci	11 (6.59%)	Cytoreduction	Maximal (no visible%)	119 (71.26%)
	Diffuse, military	36 (21.56%)		Optimal (<1 cm%)	36 (21.56%)
Small bowel serosal invasion	No	127 (76.05%)		Suboptimal (≥1 cm%)	12 (7.19%)
	Localized foci	18 (10.78%)	Intestinal resection	No	128 (76.65%)
	Diffuse, military	22 (13.17%)		Yes	39 (23.35%)
Small bowel mesentery invasion	No	110 (65.87%)	Small bowel resection	No	158 (94.61%)
	Localized foci	17 (10.18%)		Yes	9 (5.39%)
	Diffuse, military	40 (23.95%)	Colorectal anastomozis	No	137 (82.04%)
				Yes	30 (17.96%)

The table continues

Table 1. Summary of the distribution of qualitative parameters (The table continues)

Pelvic peritonectomy	No	82 (49.1%)	FIGO stage	I	33 (19.76%)
	Yes	85 (50.9%)		II	16 (9.58%)
Paracolic peritonectomy	No	117 (70.06%)		III	61 (36.53%)
	Yes	50 (29.94%)		IV	57 (34.13%)
Diafragm peritonectomy	No	139 (83.23%)	LN involvement	0	70 (41.92%)
	Yes	28 (16.77%)		1	40 (23.95%)
Splenectomy and or distal pancreatectomy	No	147 (88.02%)		2	57 (34.13%)
	Yes	20 (11.98%)	Adjuvant therapy	No	18 (10.78%)
Lymphadenectomy	No	55 (32.93%)		Yes	142 (85.03%)
	Selective LN debulking	5 (2.99%)		Lost followup	0 (0%)
	Systemic pelvic-paraaortic	107 (64.07%)		Death before adjuvant therapy	7 (4.19%)
Intraop. complication	No	150 (89.82%)	Total no of first line chemotherapy cycles ¥	0	12 (7.41%)
	Yes	17 (10.18%)		1-6 cycles of chemotherapy	93 (57.41%)
Intraop need for bloood transfusion	No	79 (47.31%)		7-8 cycles of chemotherapy	49 (30.25%)
	Yes	88 (52.69%)		Death before adjuvant therapy.lost to follow up	8 (4.94%)
Needfor ICU	No	68 (40.72%)	Disease status on or after primary therapy	Complete response	129 (77.25%)
	Yes	99 (59.28%)		Partial response	22 (13.17%)
Postoperative any adverse event including deaths	No	86 (51.5%)		Stable disease	5 (2.99%)
	Yes	81 (48.5%)		Progression	4 (2.4%)
Clavien Dindo classification of surgical advers events	No	84 (50.3%)		Death before completion of primary therapy	7 (4.19%)
	Grade 1	19 (11.38%)		Lost to follow up	0 (0%)
	Grade 2	31 (18.56%)			
	Grade 3	17 (10.18%)			
	Grade 4	10 (5.99%)			
	Grade 5	6 (3.59%)			
Tumor histotype	High grade	125 (74.9%)			
	Others	42 (25.1%)			

¥ Data is missing for five patients. The distribution percentages were calculated based on a sample of 162 patients. ECOG PS: Eastern Cooperative Oncology Group Performance Status, ICU: Intensive care unit, NACT: Neoadjuvant chemotherapy

Tablo 2. General distribution patterns of the quantitative attributes used in the ML models

Parameter	Minimum	Maximum	Median	Mean	SD
Age	22	82	57	58	11
Serum CA-125	3.4	25801.0	614.0	1548.4	3436.0
LOSH-ICU	0	38	1	2	4
LOSH (overall)	1	77	11	14	9
No. of pelvic LNs removed	12	69	29	31	12
No. of paraaortic LNs removed	8	73	26	27	13
No. of total LNs removed	15	129	56	58	19
Time interval between debulking surgery and adjuvant therapy, days	14	99	36	38	14
Recurrence time, months	22	82	57	58	11

LOSH-ICU: Length Of stay hospital ıntensive care unit, LN: Lymph node

a MAE of 16.45, a RMSE of 20.98, and a model-building time of only 0.03 seconds, thereby emerging as the most effective algorithm and warranting further investigation. In comparison, Linear Regression, Bootstrap Aggregating Additive Regression, Random Committee, and Regression by Discretization yielded CC of 0.5326, 0.4915, 0.4491, 0.4077, and 0.3817, respectively. Correlation analysis between actual and predicted PFS values for Random Forest indicated a moderate correlation (r=0.4–0.6), suggesting moderate predictive accuracy. The ability of Random Forest to reduce overfitting by combining multiple models and capture complex, non-linear relationships between features is the reason for its strong performance.[13,14] Recent studies have shown that using radiomic and multi-omic data alongside clinical information improves the accuracy of ovarian cancer predictions. For example, Jian et al.[15] developed a Random Forest model that combined imaging data with clinical information and achieved a 77.2% AUC, outperforming models based solely on clinical

Table 3. Evaluation of ML model performance and prediction capabilities for algorithms used

		Evaluation of model performance		
ML algorithms	**Time taken to build model**	**Mean absolute error**	**Root mean squared error**	**Correlation coefficient†**
Random Forest	0.03 sec	16.45	20.98	0.5731
Multilayer perceptron	0.02 sec	27.30	33.70	0.1543
Linear regression	0.02 sec	18.16	21.47	0.5326
Support vector regression	0.03 sec	25.25	29.62	0.1997
Additive regression	0.02 sec	21.47	25.83	0.4491
Bootstrap aggregating (Bagging)	0.03 sec	17.46	21.67	0.4915
Random Committee	0.03 sec	18.5026	24.0229	0.4077
Regression by discretization	0.08 sec	18.5396	26.6757	0.3817

*Training and validation split: %75 and %25, †Correlation between the actual and the predicted PFS data, ML: Machine learning, PFS: Progression-free survival

Table 4. Evaluating the correlation between the actual and predicted progression-free survival (PFS) values, as well as exploring differences within the PFS values

	Correlation*		Difference within PFS values**	
	Coefficient (r)	**p value**	**t**	**p value**
Actual PFS Predicted PFS	0.573	<0.001	-0.068	0.946

*Pearson's correlation analysis **Paired sample t-test. PFS: Progression-free survival (months)

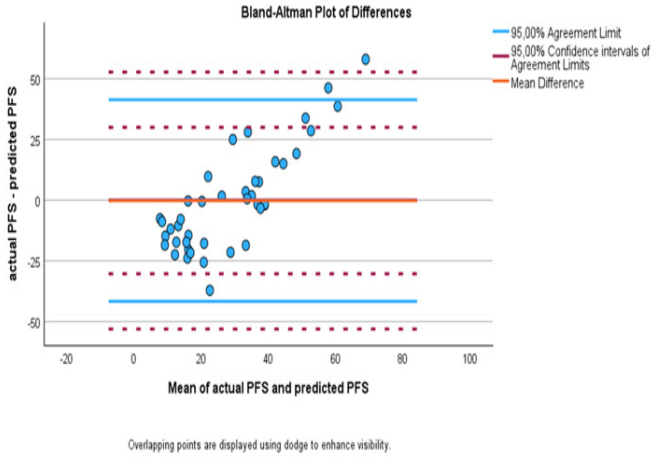


Figure 4. Bland–Altman plot of differences between actual and predicted PFS values

data. Similarly, Laios et al.[16] used ML in advanced high grade serous ovarian cancer, emphasised the importance of feature selection, and predicted 2 year survival with approximately 73% accuracy using support vector machines and ensemble models.These results demonstrate that radiomic features and accurate feature selection provide added value beyond traditional clinical methods. Furthermore, systematic reviews highlight the importance of combining different types of data. For example, Piedimonte et al.[17] reported that radiomics based ML models achieved AUC values ranging from 0.77 to 0.93 in various studies. Radiogenomic models, which combine imaging data with molecular profiles, have yielded even more promising results. A meta-analysis by Maiorano et al.[18] revealed that the AUC value of these models can reach up to 0.975, while Zeng et al.[19] deep learning-based model, which combines multi-centre imaging and genetic data, achieved the highest reported prediction accuracy (AUC=0.975).

Beyond imaging methods, multi-omic-based models have also made significant contributions. Wu et al.[20] developed the AI assisted prognostic index AIDPI by integrating transcriptomic and clinical data, demonstrating that this index improves patient risk classification. Similarly, Chen et al.[21] introduced the CSOARG model based on eight gene expression signatures, which achieved an AUC value of 0.68 in five-year survival prediction. These gene-based signatures provide valuable insights into biological risk profiles, treatment responses, and the tumours immune microenvironment. Jiang et al.[22] contributed to the field with AUTOSurv, an interpretable deep learning based platform that combines clinical, gene expression, and miRNA data, reporting that this platform outperforms traditional ML methods. Overall, these studies demonstrate that analysing multidimensional data with deep neural networks can provide accuracy beyond classical methods.

In this context, our study highlights the limitations of models based solely on clinical parameters. Using the Random Forest algorithm, a statistically significant but moderate correlation (Pearson r=0.573, p<0.001) was obtained between actual and predicted PFS. Furthermore, the absence of a significant difference between predicted and observed PFS values (p=0.946) supports the model's consistency. Although the findings demonstrate the predictive power of our ML model based on clinical data, they also indicate that its performance is limited compared to radiomic and multi-omic approaches that better reflect tumour heterogeneity and treatment response.

Nevertheless, the main strength of our study lies in its practicality and accessibility. Since the model was developed solely on the basis of routinely collected clinical data, it does not rely on advanced radiomic analyses or expensive genomic technologies. This makes it especially valuable for resource-constrained settings, where the implementation of high-performance radiomic or multi-omic models may not be feasible. Thus, although integrative models can achieve higher accuracy, a clinically based model with acceptable predictive performance can still function as a rapid, cost-effective, and complementary decision-support tool in the management of epithelial ovarian cancer. In epithelial ovarian cancer, achieving R0 resection is recognized as a critical determinant of patient survival outcomes. A comprehensive study of 571

patients sought to develop an AI-based prediction model focused on estimating the probability of R0 resection. Using the eXtreme Gradient Boosting (XGBoost) algorithm, the model incorporated multiple variables primarily related to patient characteristics and surgical features. To enhance interpretability, SHAP were applied, enabling both global and local explanations of model predictions. The XGBoost framework demonstrated strong predictive accuracy, with an AUC of 0.866 (95% confidence interval [CI]: 0.80–0.93).[23]

Cox proportional hazards regression, a commonly used method in survival analysis, is a powerful tool for evaluating the effects of covariates on the hazard function. In clinical practice, tumour characteristics, patient demographics, and treatment responses often exhibit complex, non-linear relationships with survival outcomes. Researchers, aware of these limitations, are increasingly turning to alternative approaches that can capture such dynamics. ML-based methods offer the opportunity to model these non-linear interactions more effectively, revealing patterns and relationships that traditional techniques may overlook. The application of these methods holds promise for improving prognosis accuracy, enhancing patient outcomes, and supporting more informed clinical decision-making.[24,25]

Limitations

This study has some limitations. It was conducted in a single centre with a relatively small patient group, so larger multicentre studies are needed to confirm our results. Many advanced models in the literature show high performance, but these are often based on retrospective data and internal validation.[26] For safe use in clinical practice, prospective and external validation, as well as studies on their impact on clinical decision-making, are required. In the future, adding radiomic and molecular biomarkers to clinical models and applying better feature selection methods may improve prognostic accuracy. Our findings highlight the gap between simple clinical models and advanced approaches, and suggest that hybrid models combining both may provide a good balance between accuracy and practicality in ovarian cancer prognosis. Despite the creation of a forward-looking dataset, a limitation of our retrospective design is the relatively modest sample size, which may limit the generalisability of the findings. Despite the limitations mentioned above, the value of the study lies in the careful examination of the parameters.

CONCLUSION

As a result, this study demonstrated that the Random Forest algorithm yielded better prediction results than the other methods tested and was effective in processing complex clinical data. The CC (r=0.57) indicates moderate accuracy, but this still has clinical value. In oncology, even models with moderate accuracy can assist by classifying patients according to their risk of recurrence, guiding follow up programmes, and identifying patients who may benefit from early treatment or clinical trials. The model does not replace clinical judgement but can support decision-making in multidisciplinary care. Future studies should test larger patient cohorts, refine parameters, improve data processing, and combine clinical data with radiomic or genomic information to enhance accuracy and generalisability across different healthcare settings.

ETHICAL DECLARATIONS

Ethics Committee Approval

This retrospective analysis was approved by the Scientific Ethics Committee for Medical Researches at Antalya Training and Research Hospital (Date: 07.11.2024, Decision No: 17/2).

Informed Consent

Because the study was designed retrospectively, no written informed consent form was obtained from patients.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin.* 2021;71(3):209-249. doi:10.3322/caac.21660
2. Shachar E, Raz Y, Rotkop G, et al. Cytoreductive surgery in advanced epithelial ovarian cancer: a real-world analysis guided by clinical variables, homologous recombination, and BRCA status. *Int J Gynecol Cancer.* 2025;35(6):101809. doi:10.1016/j.ijgc.2025.101809
3. Caruso G, Tomao F, Parma G, et al. Poly (ADP-ribose) polymerase inhibitors (PARPi) in ovarian cancer: lessons learned and future directions. *Int J Gynecol Cancer.* 2023;33(4):431-443. doi:10.1136/ijgc-2022-004149
4. Atallah GA, Kampan NC, Chew KT, et al. Predicting prognosis and platinum resistance in ovarian cancer: role of immunohistochemistry biomarkers. *Int J Mol Sci.* 2023;24(3):1973. doi:10.3390/ijms24031973
5. Cruz JA, Wishart DS. Applications of machine learning in cancer prediction and prognosis. *Cancer Inform.* 2006;2:117693510600200030. doi:10.1177/117693510600200030
6. Fountzilas E, Pearce T, Baysal MA, Chakraborty A, Tsimberidou AM. Convergence of evolving artificial intelligence and machine learning techniques in precision oncology. *NPJ Digit Med.* 2025;8(1):75. doi:10.1038/s41746-025-01471-y
7. Abbas S, Asif M, Rehman A, Alharbi M, Khan MA, Elmitwally N. Emerging research trends in artificial intelligence for cancer diagnostic systems: a comprehensive review. *Heliyon.* 2024;10(17):e36743. doi:10.1016/j.heliyon.2024.e36743
8. Enshaei A, Robson C, Edmondson R. Artificial intelligence systems as prognostic and predictive tools in ovarian cancer. *Ann Surg Oncol.* 2015;22:3970-3975. doi:10.1245/s10434-015-4475-6
9. Krzyszczyk P, Acevedo A, Davidoff EJ, et al. The growing role of precision and personalized medicine for cancer treatment. *Technol Cancer Res Treat.* 2018;6(03n04):79-100. doi:10.1142/S2339547818300020
10. Fagotti A, Ferrandina MG, Vizzielli G, et al. Randomized trial of primary debulking surgery versus neoadjuvant chemotherapy for advanced epithelial ovarian cancer (SCORPION-NCT01461850). *Int J Gynecol Cancer.* 2020;30(11):1657-1664. doi:10.1136/ijgc-2020-001640

11. Van Meurs HS, Tajik P, Hof MH, et al. Which patients benefit most from primary surgery or neoadjuvant chemotherapy in stage IIIC or IV ovarian cancer? An exploratory analysis of the European Organisation for Research and Treatment of Cancer 55971 randomised trial. *Eur J Cancer*. 2013;49(15):3191-3201. doi:10.1016/j.ejca.2013.06.013
12. Clavien PA, Barkun J, De Oliveira ML, et al. The Clavien-Dindo classification of surgical complications: five-year experience. *Ann Surg*. 2009;250(2):187-196. doi:10.1097/SLA.0b013e3181b13ca2
13. Biau G, Scornet E. A random forest guided tour. *Test*. 2016;25(2):197-227.
14. Fernández-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? *J Machine Learning Res*. 2014;15(1):3133-3181.
15. Jian L, Chen X, Hu P, et al. Predicting progression-free survival in patients with epithelial ovarian cancer using an interpretable random forest model. *Heliyon*. 2024;10(15):e35344. doi:10.1016/j.heliyon.2024.e35344
16. Laios A, Katsenou A, Tan YS, et al. Feature selection is critical for 2-year prognosis in advanced stage high grade serous ovarian cancer by using machine learning. *Cancer Control*. 2021;28:10732748211044678. doi:10.1177/10732748211044678
17. Piedimonte G, Fusco R, Rizzo A, et al. Machine learning models based on radiomics for the prediction of ovarian cancer outcomes: a systematic review. *Cancers (Basel)*. 2025;17(3):336. doi:10.3390/cancers17030336
18. Maiorano BA, Rabaiotti E. Prognostic and predictive potential of radiogenomics in epithelial ovarian cancer: a systematic review and meta-analysis. *Onco*. 2025;6(4):84. doi:10.3390/onco6040084
19. Zeng S, Wang XL, Yang H. Radiomics and radiogenomics: extracting more information from medical images for the diagnosis and prognostic prediction of ovarian cancer. *Mil Med Res*. 2024;11(1):77. doi:10.1186/s40779-024-00580-1
20. Wu J, Zhao L, Li Y, et al. AIDPI: a transcriptomic and multi-omics prognostic model for ovarian cancer. *Front Genet*. 2025;16:12268085.
21. Chen X, Liu Y, Zhang H, et al. Construction and validation of a CSOARG gene signature for prognosis in ovarian cancer. *Front Oncol*. 2025;15:1592426. doi:10.3389/fonc.2025.1592426
22. Jiang Y, Yang C, Xu W, et al. AUTOSurv: deep learning framework integrating multi-omics and clinical data for ovarian cancer survival prediction. *NPJ Precis Oncol*. 2024;8:37. doi:10.1038/s41698-023-00494-6
23. Laios A, Kalampokis E, Johnson R, et al. Explainable artificial intelligence for prediction of complete surgical cytoreduction in advanced-stage epithelial ovarian cancer. *J Pers Med*. 2022;12(4):607. doi:10.3390/jpm12040607
24. Harrell J, Frank E, Harrell FE. Cox proportional hazards regression model. In: Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis. 2nd ed. Springer; 2015. doi:10.1007/978-3-319-19425-7
25. Bertolaccini L, Pardolesi A, Davoli F, Solli P. Nanos gigantium humeris insidentes: the awarded Cox proportional hazards model. *J Thorac Dis*. 2016;8(11):3464-3470. doi:10.21037/jtd.2016.11.84
26. Jennings B, Zhang T. A systematic review of multimodal AI models for ovarian cancer prognosis using whole slide imaging and omics data. *arXiv preprint*. 2025;arXiv:2507.16876.

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1746227

Anatolian Curr Med J. 2025;7(5):695-702

Fabricated or accurate? Ethical concerns and citation hallucination in AI-generated scientific writing on musculoskeletal topics

Ertuğrul Safran[1], Adem Çalı[2]

[1]*Department of Physiotherapy and Rehabilitation, Faculty of Health Sciences, Bezmiâlem Vakıf University, İstanbul, Turkiye*
[2]*Department of Occupational Therapy, Faculty of Health Sciences, İstanbul Topkapı University, İstanbul, Turkiye*

Cite this article as: Safran E, Çalı A. Fabricated or accurate? Ethical concerns and citation hallucination in AI-generated scientific writing on musculoskeletal topics. *Anatolian Curr Med J.* 2025;7(5):695-702.

Received: 19.07.2025 ◆ **Accepted**: 10.09.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: Large language models (LLMs) such as ChatGPT are increasingly used in academic and clinical writing. While these tools can generate coherent and domain-specific text, concerns persist regarding the accuracy of their automatically generated references. In musculoskeletal rehabilitation—a field heavily reliant on current evidence—the reliability of citations is especially critical. Yet, systematic evaluations of citation accuracy in AI-generated scientific content are lacking. To evaluate the reference accuracy of scientific texts generated by ChatGPT (GPT-4) in response to musculoskeletal rehabilitation prompts, and to determine whether reference accuracy improves following structured post-generation verification.

Methods: ChatGPT was prompted to generate four scientific paragraphs on musculoskeletal rehabilitation topics (manual therapy, ACL reconstruction, low back pain, and rotator cuff repair), each including 10 references with DOIs. A total of 40 references were analyzed using a 3-point scoring system (0=fabricated, 1=partially correct, 2=fully accurate), which was used to assess citation quality. After initial evaluation, ChatGPT was asked to verify and revise its references. Scores before and after this step were compared descriptively and with Wilcoxon signed-rank tests to assess statistical significance, and effect sizes (r) were calculated to estimate the magnitude of improvement.

Results: Only 7.5% of references were fully accurate in the initial generation, while 42.5% were completely fabricated. The remaining 50% were partially correct. After verification, the proportion of fully accurate references rose to 77.5%. Wilcoxon signed-rank testing confirmed a statistically significant improvement in accuracy across all prompts ($W=561.0$, $p<0.001$, $r=0.60$). The most common errors included invalid DOIs, fabricated article titles, and mismatched metadata.

Conclusion: ChatGPT can generate coherent scientific content, but its initial references are frequently inaccurate or fabricated. Structured post-generation verification significantly improves reference accuracy, as confirmed by statistical testing. These findings suggest that LLMs may be integrated as drafting tools in academic and clinical musculoskeletal contexts, but only when accompanied by strict human-led verification of citations.

Keywords: ChatGPT, artificial intelligence, musculoskeletal rehabilitation, scientific writing, reference accuracy, hallucinated citations

INTRODUCTION

The integration of artificial intelligence (AI) into academic and scientific writing has introduced powerful tools for generating domain-specific content with unprecedented speed. Among these technologies, large language models (LLMs) such as ChatGPT have demonstrated remarkable fluency in producing coherent and technically structured scientific texts across various disciplines, including health sciences and rehabilitation domains.[1-4] However, a critical and underexplored concern lies in the accuracy and reliability of references generated by these models. In musculoskeletal rehabilitation—including physiotherapy, exercise science, and manual therapy—clinicians, researchers, and educators rely heavily on accurate and up-to-date scientific literature to guide evidence-based decision-making. Citation accuracy is therefore not merely a technical detail but a cornerstone of trustworthy clinical, educational, and research practice. When references generated by LLMs are inaccurate or fabricated, the consequences extend beyond academic integrity, posing risks for patient safety, clinical outcomes, and professional training.[5-8]

Recent investigations have increasingly drawn attention to the phenomenon of "hallucinated references," wherein LLMs such as ChatGPT generate citations that appear syntactically correct and contextually appropriate, yet do not correspond to any real or retrievable publication.[9-12] These fabricated references may include convincing combinations

Corresponding Author: Ertuğrul Safran, esafran@bezmialem.edu.tr

of real author names, journal titles, and plausible publication years, making them difficult to detect without thorough verification.[13] This issue poses a significant challenge in health sciences, where evidence-based accuracy is foundational, and its consequences become particularly acute in clinical fields. Misinformation introduced by inaccurate or fabricated references may propagate unnoticed through educational resources, systematic reviews, or even AI-supported clinical tools—potentially influencing clinical recommendations and practice standards.[14,15] While the growing use of AI-based writing tools introduces opportunities for increased efficiency, it also carries inherent risks associated with unverified bibliographic output.[16-19] For example, an erroneous reference regarding the effectiveness of manual therapy could lead to inappropriate treatment planning, while a fabricated citation on exercise prescription might expose patients to ineffective or unsafe interventions. These risks highlight the necessity of rigorous verification when integrating AI-generated content into musculoskeletal rehabilitation research and practice.

In contrast to previous anecdotal reports or general evaluations,[20,21] there remains a lack of systematic research quantifying the accuracy and correction potential of hallucinated references in musculoskeletal-focused scientific content. Addressing this gap is essential to ensuring the responsible integration of LLMs into scholarly and clinical communication. Therefore, the aim of this study was to systematically evaluate the accuracy of references generated by ChatGPT in the context of musculoskeletal rehabilitation and to examine whether post-generation prompting could improve citation reliability.

METHODS

Ethics

This study did not involve human or animal participants. Therefore, ethical approval was not required according to national guidelines. All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Study Design

This was a structured observational analysis comparing four prompt conditions used to generate scientific paragraphs and associated reference lists using ChatGPT (GPT-4 model). All prompts were designed to cover topics related to musculoskeletal disorders in the context of rehabilitation and therapeutic strategies.

Prompt Conditions

The prompts provided to ChatGPT included a consistent instruction to generate a scientific paragraph (approximately 300–400 words) along with 10 references including DOIs.

The following prompts—each of which addressed a clinical topic within the musculoskeletal rehabilitation field—were used:

Prompt 1: Write a short scientific text (approximately 300–400 words) on effects of manual therapy on pain, including 10 references with DOIs.

Prompt 2: Write a short scientific text (approximately 300–400 words) on ACL reconstruction rehabilitation strategies, including 10 references with DOIs.

Prompt 3: Write a short scientific text (approximately 300–400 words) on low back pain exercise strategies, including 10 references with DOIs.

Prompt 4: Write a short scientific text (approximately 300–400 words) on rotator cuff repair rehabilitation strategies, including 10 references with DOIs.

The four clinical topics were deliberately chosen to reflect both breadth and clinical relevance within musculoskeletal rehabilitation. One topic (manual therapy and pain) represents a broad theme that is central across many rehabilitation domains, while the others were selected to cover frequently encountered conditions at different anatomical levels: lower extremity (ACL reconstruction), upper extremity (rotator cuff repair), and spine (low back pain). This distribution allowed for a representative sampling of musculoskeletal rehabilitation contexts without narrowing the analysis to a single region. The instruction to generate 10 references per prompt was used to standardize outputs and to reflect a realistic academic scenario, as users commonly request short scientific texts accompanied by a concise list of approximately ten references. This limit also ensured that the reference lists were manageable for systematic verification while minimizing variability due to overly extensive outputs.

The prompts were administered sequentially in the fixed order listed above (Prompt 1 through Prompt 4). The sequence was not randomized, as the study design aimed to compare outcomes across consistent and replicable conditions. Each prompt was entered in a separate, context-cleared session to ensure that ChatGPT's outputs were not influenced by prior responses.

Following the generation of each response, ChatGPT was given an additional instruction to review the same list of references it had just generated. The exact wording of this post-verification prompt was: 'Please check the references and DOIs you have just generated. Indicate whether each reference exists, whether the metadata (authors, title, journal, year) is correct, and whether the DOI is functional. Correct any errors if possible.' This prompt did not request a completely new reference list; rather, it asked the model to re-examine and, where possible, correct the citations it had already provided. This procedure aimed to simulate a self-correction process and to assess whether the model could improve reference accuracy upon reevaluation. All outputs were generated using ChatGPT (GPT-4, OpenAI web interface, accessed in June 2025). At the time of data collection, OpenAI did not provide a publicly available version number beyond the model designation (GPT-4); therefore, the model is reported with its platform and date of access for transparency.

Reference Evaluation

Each generated reference was evaluated based on the following criteria:

Existence: Does the reference exist in the scientific literature?

Correct author(s): Are the cited authors correct?

Correct title: Is the article title accurate?

Correct journal information: Are the journal name, volume, issue, and pages accurate?

Valid DOI: Does the DOI link resolve to the correct article?

A scoring system was used:

0=Fabricated reference or completely mismatched information

1=Partially correct (e.g., wrong journal or year, incorrect DOI)

2=Fully accurate (exact match across all fields)

In this study, one researcher was responsible for prompt design and content generation, while another independently evaluated the accuracy of the references and DOIs. Both researchers had over 10 years of professional experience in musculoskeletal rehabilitation and academic writing. When DOIs provided by ChatGPT did not resolve automatically, references were verified manually using Crossref, PubMed, Google Scholar, and publisher websites. This predefined protocol ensured that fabricated or mismatched DOIs could be identified and corrected when possible.

Sample Size and Statistical Analysis

A total of 40 references were analyzed (10 references per prompt×4 prompt conditions). Descriptive statistics were used to compare the accuracy scores between prompt conditions. The frequency of score categories (0, 1, 2) was presented as percentages. Additionally, error types were qualitatively categorized (e.g., fabricated reference, wrong DOI, mismatched journal). To further assess changes in reference accuracy before and after verification, Wilcoxon signed-rank tests were performed, with effect sizes (r) calculated to indicate the magnitude of improvement. Statistical significance was set at $p<0.05$. All analyses were conducted using IBM SPSS Statistics 25.

RESULTS

Prompt 1: Effects of Manual Therapy on Pain

A total of 10 references generated in response to prompt 1 were evaluated. None of the references received a full score of 2. The majority of the references (8 out of 10) were partially accurate, receiving a score of 1, typically due to incorrect or missing DOIs, publication years, or mismatched journal details. Two references (20%) were entirely fabricated or had major mismatches (score=0). A detailed breakdown of each reference and its specific issues is presented in **Table 1**.

After prompting ChatGPT to review and verify the accuracy of the references and DOIs, a substantial improvement was observed. Seven references (70%) received a fully accurate score of 2 upon regeneration and verification. Two references (20%) remained partially correct (score=1), and one reference (10%) was still found to be fabricated or completely mismatched (score=0).

Table 1. Initial accuracy evaluation of references generated by prompt 1 (manual therapy on pain)

Reference	Score	Explanation
1	1	DOI and year are incorrect
2	0	Same title exists but by different authors; DOI is non-functional
3	1	DOI and year are incorrect
4	1	Author and topic are similar, but title and journal are different
5	1	DOI is incorrect; author list partially matches a real article
6	1	DOI and year are incorrect
7	1	DOI non-functional; journal name incorrect
8	1	DOI missing; all other citation details are correct
9	0	Fabricated title and DOI under a real author's name
10	1	Correct title and author; journal name and DOI incorrect

DOI: Digital object identifier. DOI validation was conducted manually using Crossref and PubMed, with supplementary checks through Google Scholar when necessary (accessed June 2025).

Prompt 2: ACL Reconstruction Rehabilitation Strategies

Ten references generated in response to prompt 2 were evaluated. None received a fully accurate score of 2 on the initial evaluation. Five references (50%) were partially accurate (score=1), typically due to incorrect DOIs or mismatched journal details, although the articles themselves existed and were relevant to the topic. The remaining five references (50%) were assigned a score of 0, either because they included fabricated article titles, non-functional DOIs, or mismatched information under real author names.

Fabrication was a recurring issue: multiple references contained fabricated article names attributed to real authors who had published similar work in the same year. These cases suggest a pattern in which the AI model tends to blend existing authorship with invented or misattributed titles. A detailed breakdown of each reference is presented in **Table 2**.

After instructing ChatGPT to verify and correct the references and DOIs, the accuracy markedly improved. Nine references (90%) received a full accuracy score (score=2), while one reference (10%) remained partially accurate (score=1).

Prompt 3: Low Back Pain Exercise Strategies

Ten references generated in response to prompt 3 were evaluated. Two references (20%) were fully accurate (score=2), while four references (40%) were partially accurate (score=1), and four references (40%) were completely inaccurate or fabricated (score=0). Inaccuracies frequently involved fabricated article titles linked to real authors, or incorrect DOIs. In some cases, the author, year, and journal details were correct, but the title was either abbreviated or artificially constructed. A detailed breakdown of each reference is presented in **Table 3**.

After asking ChatGPT to verify and correct the references and DOIs, a clear improvement in accuracy was observed. Eight references (80%) achieved a fully accurate score (score=2),

Table 2. Initial accuracy evaluation of references generated by prompt 2 (ACL reconstruction rehabilitation strategies)

Reference	Score	Explanation
1	1	Article exists; correct metadata but DOI incorrect and refers to a different ACL-related article
2	1	Article exists; journal name is incorrect, DOI incorrect and refers to a different ACL-related article
3	0	Invalid DOI; fabricated article title under real author who has similar publications
4	0	Invalid DOI; fabricated article title, though the same author published a similar topic in the same year
5	1	Article exists; correct information but DOI is invalid; requires manual search
6	1	Article exists; DOI incorrect and refers to a different ACL-related paper
7	0	Incorrect DOI; fabricated title under real author who has similar work
8	0	Invalid DOI; fabricated title under real author with similar publications
9	0	DOI functional; fabricated title; author and year are correct; journal name incorrect
10	1	DOI works; article content is correct but the title is shortened

DOI: Digital object identifier. DOI validation was conducted manually using Crossref and PubMed, with supplementary checks through Google Scholar when necessary (accessed June 2025).

Table 3. Initial accuracy evaluation of references generated by prompt 3 (low back pain exercise strategies)

Reference	Score	Explanation
1	2	Fully accurate
2	2	Fully accurate
3	0	Invalid DOI; fabricated title under real author with related work
4	1	DOI correct; author, year, journal correct; title closely matches but fabricated
5	1	DOI correct; metadata accurate; title similar but not real
6	0	DOI incorrect; fabricated title; author has related publication
7	0	Invalid DOI; fabricated title under real author with similar topic
8	1	DOI and metadata accurate; title abbreviated but acceptable
9	0	Invalid DOI; fabricated title under real author with similar topic
10	1	DOI incorrect; all other citation details correct but required manual search

DOI: Digital object identifier. DOI validation was conducted manually using Crossref and PubMed, with supplementary checks through Google Scholar when necessary (accessed June 2025).

while two references (20%) remained partially accurate (score=1). No references received a score of 0 following this verification step, suggesting that secondary prompting significantly enhances citation precision.

Prompt 4: Rotator Cuff Repair Rehabilitation Strategies

Ten references generated in response to prompt 4 were evaluated. Only one reference (10%) was fully accurate, while three (30%) were partially accurate. Six (60%) references contained fabricated titles, incorrect DOIs, or misattributions. Among these, 4/10 DOIs did not resolve, 3/10 required manual correction via external searches, and 3/10 remained unverifiable. The remaining six references (60%) were assigned a score of 0, as they included fabricated article titles, incorrect DOIs, or misattributions to authors who had published similar but unrelated works.

This prompt resulted in the lowest initial accuracy among all four prompts, with a high rate of fabricated citations and non-functional DOIs. The fabricated content often used real authors but invented titles or incorrectly matched publications. Details are summarized in Table 4.

Table 4. Initial accuracy evaluation of references generated by prompt 4 (rotator cuff repair rehabilitation strategies)

Reference	Score	Explanation
1	0	DOI incorrect; fabricated title under real author with similar topic
2	2	Fully accurate
3	0	DOI incorrect; fabricated title under real author with different paper
4	1	Article exists; DOI invalid; manual search required
5	1	Article exists; DOI invalid; manual search required
6	1	Article exists; DOI invalid; manual search required
7	0	Invalid DOI; real author, but incorrect publication
8	0	Invalid DOI; fabricated title under real author
9	0	DOI incorrect; fabricated article title under real author
10	0	Invalid DOI; fabricated metadata

DOI: Digital object identifier. DOI validation was conducted manually using Crossref and PubMed, with supplementary checks through Google Scholar when necessary (accessed June 2025).

Following a verification prompt to ChatGPT, seven references (70%) were corrected to full accuracy (score=2), while three references (30%) remained either partially correct (score=1) or unverifiable. This again demonstrates an improvement after issuing a follow-up prompt focused on citation verification.

Out of 40 references, only 3 (7.5%) were fully accurate (score=2) without any correction. A large proportion, 20 references (50%), were partially accurate (score=1), while 17 references (42.5%) were found to be completely fabricated or contained major errors (score=0).

The post-verification prompt significantly improved accuracy. In total, 31 out of 40 references (77.5%) were corrected to full accuracy (score=2) after ChatGPT was asked to verify and revise its citations.

In addition to descriptive reporting, inferential analyses were conducted to evaluate changes in reference accuracy before and after the verification step. Across all 40 references, median accuracy increased from 1 (IQR 0–1) before verification to 2 (IQR 2–2) after verification. The proportion of fully accurate references also rose substantially, from 7.5% (3/40) to 77.5% (31/40). A Wilcoxon signed-rank test confirmed that this improvement was statistically significant ($W=561.0$, $p<0.001$, $r=0.60$), representing a large effect size.

Prompt-level analyses similarly demonstrated significant improvements in accuracy:

Prompt 1 (manual therapy on pain): median score increased from 1 to 2 (W=32.5, p=0.017).

Prompt 2 (ACL reconstruction rehabilitation): median score increased from 0 to 2 (W=45.0, p=0.003).

Prompt 3 (low back pain exercise): median score increased from 1 to 2 (W=48.0, p=0.019).

Prompt 4 (rotator cuff repair rehabilitation): median score increased from 0 to 2 (W=34.5, p=0.008).

These findings indicate that the verification prompt produced consistent and statistically significant improvements in citation accuracy across all tested topics.

A detailed cross-prompt comparison of reference accuracy before and after verification is presented in Table 5.

Table 5. Summary of reference accuracy scores across all prompts initial and post verification

Prompt topic	Score=0	Score=1	Score=2	After verification (score=2)	Wilcoxon p value
Prompt 1–manual therapy on pain	2	8	0	7	0.017
Prompt 2–ACL reconstruction rehab	5	5	0	9	0.003
Prompt 3–low back pain exercise	4	4	2	8	0.019
Prompt 4–rotator cuff repair rehab	6	3	1	7	0.008
Total	17	20	3	31	<0.001

p: Statistical significance value

The distribution of reference accuracy scores before and after the verification step is illustrated in Figure as stacked column charts. The figure demonstrates a marked shift from fabricated (score=0) and partially correct (score=1) references toward fully accurate references (score=2) across all prompts. In the pooled dataset (n=40), the proportion of fully accurate references increased from 7.5% before verification to 77.5% after verification, while fabricated references declined from 42.5% to 2.5%.

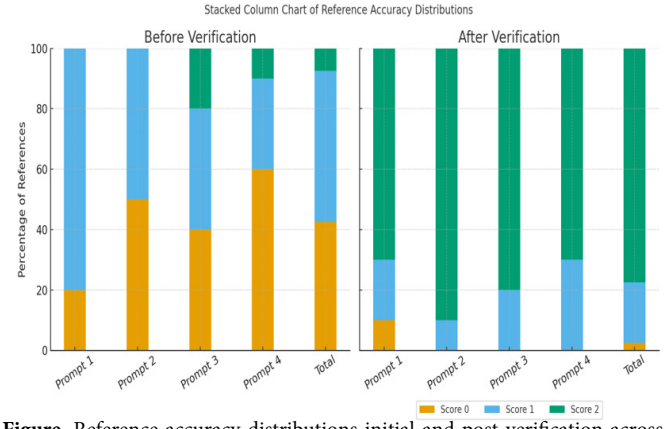


Figure. Reference accuracy distributions initial and post verification across all prompts

DISCUSSION

This study systematically investigated the accuracy of references generated by ChatGPT in response to four distinct prompts related to musculoskeletal rehabilitation. Our findings revealed a high rate of citation errors in the model's initial outputs, with only 7.5% of references being fully accurate and more than 40% being entirely fabricated. The most frequent issues included non-functional or incorrect DOIs, fabricated article titles, and mismatched author or journal information—often combining real author names with invented bibliographic data. However, a marked improvement was observed when ChatGPT was prompted to verify and correct its references: the proportion of fully accurate citations increased to 77.5%. Wilcoxon signed-rank testing confirmed that this improvement was statistically significant (W=561.0, p<0.001, r=0.60), indicating a large effect size. This improvement may reflect the mechanism of self-critique prompting, whereby the model re-evaluates and refines its own outputs when explicitly asked to verify bibliographic accuracy. In our analysis, this iterative process was particularly effective in converting partially correct references into fully accurate ones, suggesting that structured follow-up prompts can substantially enhance citation reliability. These results highlight both the potential and the current limitations of LLMs in generating credible scientific content, especially in fields such as physiotherapy and rehabilitation where evidence-based information is essential.

A closer analysis of the errors across all prompts shows that fabricated references often borrowed credible author names and plausible-sounding journal titles, but the combinations did not correspond to any actual publication. This tendency to prioritize coherence over factuality could be especially problematic in health sciences, where citation accuracy directly impacts clinical, educational, and research practices. Notably, baseline accuracy was lowest for prompt 4 (rotator cuff repair), where only 10% of references were fully accurate prior to verification. Several factors may account for this finding. First, rotator cuff rehabilitation is a more specialized clinical area compared to low back pain or ACL rehabilitation, with a relatively narrower pool of highly cited publications. This may have increased the likelihood of the model fabricating plausible but non-existent references. Second, DOIs for certain articles in surgical or rehabilitation subfields are less consistently indexed across databases, further complicating automatic generation and verification. Finally, the imposed requirement of providing exactly ten references may have amplified the model's tendency to hallucinate citations when genuine sources were limited. Together, these factors likely contributed to the lower baseline accuracy for prompt 4, underscoring the importance of cautious interpretation when AI systems are tasked with generating references in narrower subdomains.

Our results align with previous research demonstrating significant challenges associated with fabricated and inaccurate references produced by ChatGPT.[22-24] Several studies have shown that while the model can generate plausible-sounding citations, many of these are either partially incorrect or entirely fabricated, often combining real author names with nonexistent

titles or invalid DOIs.[25-29] These hallucinated references have been documented across various disciplines, including medicine, education, and health sciences, raising widespread concerns about the reliability of AI-generated bibliographic content. Our findings reinforce these observations within the musculoskeletal domain, providing structured, topic-specific evidence that highlights the persistent risks of using LLMs without rigorous reference verification. Gravel et al.[20] found that ChatGPT often provides responses unsuitable for scientific publication, coupled with convincingly fabricated citations that require careful verification. Orduña-Malea and Cabezas-Clavijo[21] similarly warned about the rise of "ghost bibliographic references," urging vigilance from journals and publishers. Furthermore, Walters and Wilder[26] evaluated 636 citations generated across 42 topics and reported that 55% of GPT-3.5 citations were fabricated compared to 18% for GPT-4. Among real citations, 43% of GPT-3.5 references and 24% of GPT-4 references contained substantive errors. Although GPT-4 represents a clear improvement, these findings underscore that substantial reliability problems persist. Additionally, Frosolini et al.[30] compared 20 AI chats and reported that ChatGPT-4.0 generated 138 unique references with a median of 7 (IQR: 1.25) per prompt, of which 4.5 (IQR: 3.0) were classified as true and 2.5 (IQR: 3.25) as erroneous or inexistent (25.7%). By contrast, ChatGPT-3.5 generated 120 references with a median of 6 (IQR: 0.00), only 1 (IQR: 2.0) true and 4.5 (IQR: 2.25) erroneous or inexistent (83.3%). This performance was significantly worse than that of ChatGPT-4.0 (Fisher's exact test, $p<0.00001$), confirming a substantial improvement in reference reliability with the newer model. However, both versions still displayed a tendency to provide erroneous or non-existent citations. This highlights a broader concern regarding LLMs and their readiness for unsupervised academic use.

It is crucial to address this challenge to maintain the reliability of scientific literature. Journals and institutions should establish strategies and good-practice principles in the evolving landscape of AI-assisted scientific writing. A recent study by Giray[31] also explored this issue, demonstrating that ChatGPT-generated references for journal articles were often entirely fabricated and failed to adhere to APA 7 formatting. Another study investigating AI hallucination in ChatGPT references found that 69 out of 178 references lacked DOIs and 28 could not be located at all, reinforcing the need for cautious use in academic writing.[32] Another analysis of ChatGPT-generated medical content revealed that 47% of references were fabricated, 46% were authentic but inaccurate, and only 7% were both authentic and accurate. Common errors included incorrect PMIDs (93%), volumes, pages, and publication years. These findings highlight the high prevalence of citation inaccuracies in medical prompts and reinforce the necessity of reference verification when using AI-generated content.[33] These findings align with our results and underscore the need for structured prompting and post-generation verification. Moreover, integrating automated fact-checking or citation validation tools directly into AI writing platforms could help mitigate the risk of hallucinated references.

Importantly, this study highlights the need for human verification mechanisms when incorporating LLM-generated references into scholarly work. Without strict guidance, models like ChatGPT may introduce misleading citations that could undermine academic integrity. On the other hand, our findings also suggest that LLMs are capable of producing highly accurate references—if given clear constraints such as requiring valid DOIs and focused content domains.

Limitations and Strengths

This study has several limitations that should be acknowledged. First, the sample size was restricted to 40 references across four prompt conditions, which may limit the generalizability of the findings. Although the selected prompts were clinically relevant, they do not encompass the full scope of musculoskeletal rehabilitation. Second, the assessment of reference accuracy was conducted manually and, despite predefined scoring criteria, remains subject to interpretation bias. The verification step also relied on ChatGPT's own response to follow-up prompts, which, while reflective of real-world use, does not represent an external validation process. Finally, the study focused solely on the GPT-4 version of ChatGPT; findings may differ across other models or platforms, including GPT-3.5, Google Gemini, or future LLMs.

Despite these limitations, the study has several strengths: it used clinically relevant and structured prompts, applied a systematic scoring framework, and involved two experienced rehabilitation researchers in the evaluation process, enhancing the credibility of the findings. Moreover, the significant improvement observed after verification (77.5%) suggests that structured prompting and post-generation checking can substantially enhance reference reliability, pointing to feasible strategies for future model optimization.

Future Research Directions

Future studies should expand the scope of analysis to include a broader range of clinical topics within musculoskeletal rehabilitation and beyond, in order to assess whether reference accuracy varies by subdomain complexity or publication density. Including a larger sample of prompts and references would enhance the generalizability of results and allow for statistical comparisons between prompt types, clinical areas, or different LLM platforms. Comparative evaluations across various LLMs —such as GPT-3.5, Google Gemini, or open-source alternatives—could also help identify performance differences and inform best practice.

Importantly, this line of research also highlights the future potential of LLMs. The significant improvement observed in our study after verification demonstrates that relatively simple interventions—such as structured prompting and iterative re-prompting—can substantially increase reference accuracy. As models continue to evolve, embedding automated fact-checking and citation validation tools (e.g., Crossref or PubMed integration) directly into AI platforms could further minimize hallucinations. With these advancements, LLMs may become reliable collaborators in academic writing,

accelerating literature synthesis while safeguarding scientific integrity.

Practical Recommendations and Improvement Strategies

Beyond identifying the challenges, it is equally important to outline practical strategies that may improve the reliability of AI-assisted academic writing. One immediate step is the integration of DOI-verification tools such as Crossref or PubMed APIs into editorial and peer-review workflows, ensuring that fabricated or mismatched references are detected early. Automated reference-screening software could also be embedded directly into manuscript submission systems, enabling authors and editors to flag inaccurate citations before publication. On the user side, researchers and students should be trained to critically evaluate AI-generated bibliographies, combining manual checks with systematic verification protocols. Finally, the development of advanced AI models with built-in fact-checking and citation validation functions represents a promising future direction that could significantly enhance the reliability of LLMs in academic and clinical contexts.

Clinical and Educational Implications

The findings of this study have direct relevance for clinical and educational practices within the field of musculoskeletal rehabilitation. In domains such as physiotherapy, manual therapy, and exercise-based rehabilitation, practitioners and educators rely heavily on current literature to guide evidence-based decision-making. The use of fabricated or inaccurate references in AI-generated texts poses a risk of disseminating misinformation, which could negatively influence clinical reasoning, therapeutic choices, or academic instruction. As LLMs become increasingly incorporated into musculoskeletal research and education, it is essential to adopt strategies such as structured prompt design, manual citation verification, and critical evaluation training for students and clinicians. Educators should explicitly address the limitations of AI-generated content, fostering a cautious and reflective approach among learners.

At the same time, the results of this study also point to the positive potential of LLMs in clinical and educational contexts. When supported by structured prompting and coupled with external validation tools, models like ChatGPT can serve as effective assistants in generating draft content, synthesizing literature, and supporting teaching workflows. If combined with automated fact-checking, LLMs may not only facilitate efficiency but also enhance training in evidence-based rehabilitation, offering students and clinicians an opportunity to engage critically with AI-generated material while reinforcing standards of academic integrity.

CONCLUSION

This study demonstrated that while ChatGPT can generate coherent scientific text in the context of musculoskeletal rehabilitation, its initial reference accuracy remains unreliable. A majority of the citations were either partially incorrect or entirely fabricated. However, structured follow-up prompts led to substantial improvements in citation validity. These findings suggest that LLMs may be useful in supporting early-stage scientific writing, but must be coupled with strict human-led verification for references. For clinical and academic practice, this implies that AI-assisted writing should always be paired with systematic reference verification protocols, the integration of DOI-checking tools where possible, and critical evaluation training for researchers and students. Ultimately, the responsible use of AI in academic and clinical musculoskeletal settings will depend on combining technological capabilities with human oversight to ensure scientific and clinical integrity.

ETHICAL DECLARATIONS

Ethics Committee Approval

This study did not involve human or animal participants. Therefore, ethical approval was not required according to national guidelines.

Informed Consent

Since no human or animal participants are involved in this study, informed consent is not required.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Mondal H, Mondal S. ChatGPT in academic writing: maximizing its benefits and minimizing the risks. *Indian J Ophthalmol.* 2023;71(12):3600-3606. doi:10.4103/IJO.IJO_718_23
2. Bom H-SH. Exploring the opportunities and challenges of ChatGPT in academic writing: a roundtable discussion. *Nucl Med Mol Imaging.* 2023;57(4):165-1677. doi:10.1007/s13139-023-00809-2
3. Jarrah AM, Wardat Y, Fidalgo P. Using ChatGPT in academic writing is (not) a form of plagiarism: what does the literature say. *Online J Commun Media Technol.* 2023;13(4):e202346. doi:10.30935/ojcmt/13572
4. Gruda D. Three ways ChatGPT helps me in my academic writing. *Nature.* 2024;10:1-6. doi:10.1038/d41586-024-01042-3
5. Švab I, Klemenc-Ketiš Z, Zupanič S. New challenges in scientific publications: referencing, artificial intelligence and ChatGPT. *Slov J Public Health.* 2023;62(3):109-112. doi:10.2478/sjph-2023-0015
6. Jan R. Examining the reliability of ChatGPT: identifying retracted scientific literature and ensuring accurate citations and references. In: Impacts of generative ai on the future of research and education. Hershey, PA: IGI Global; 2025.
7. Frosolini A, Gennaro P, Cascino F, Gabriele G. In reference to "role of Chat GPT in public health", to highlight the AI's incorrect reference generation. *Ann Biomed Eng.* 2023;51(10):2120-2122. doi:10.1007/s10439-023-03248-4
8. Cohen F, Vallimont J, Gelfand AA. Caution regarding fabricated citations from artificial intelligence. *Headache.* 2024;64(1):133-135. doi:10.1111/head.14649

9. Borji A. A categorical archive of ChatGPT failures. arXiv. 2023;2302.03494. doi:10.48550/arXiv.2302.03494
10. Bai Z, Wang P, Xiao T, et al. Hallucination of multimodal large language models: a survey. *arXiv.* 2024;2404.18930. doi:10.48550/arXiv.2404.18930
11. Metze K, Morandin-Reis RC, Lorand-Metze I, Florindo JB. Bibliographic research with ChatGPT may be misleading: the problem of hallucination. *J Pediatr Surg.* 2024;59(1):158-159. doi:10.1016/j.jpedsurg.2023.08.018
12. Aljamaan F, Temsah M-H, Altamimi I, et al. Reference hallucination score for medical artificial intelligence chatbots: development and usability study. *JMIR Med Inform*. 2024;12(1):e54345. doi:10.2196/54345
13. Mohammad B, Supti T, Alzubaidi M, et al. The pros and cons of using ChatGPT in medical education: a scoping review. In: Healthcare transformation with informatics and artificial intelligence. Volume 305. 2023.
14. Gao CA, Howard FM, Markov NS, et al. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. *bioRxiv.* 2022;12(23):521610. doi:10.1101/2022.12.23.521610
15. Sharun K, Banu SA, Pawde AM, et al. ChatGPT and artificial hallucinations in stem cell research: assessing the accuracy of generated references–a preliminary study. *Ann Med Surg (Lond).* 2023;85(10):5275-5278. doi:10.1097/MS9.0000000000001228
16. Safran E, Yildirim S. A cross-sectional study on ChatGPT's alignment with clinical practice guidelines in musculoskeletal rehabilitation. *BMC Musculoskelet Disord*. 2025;26(1):411. doi:10.1186/s12891-025-08650-8
17. Lancaster T. Artificial intelligence, text generation tools and ChatGPT–does digital watermarking offer a solution? *Int J Educ Integr.* 2023;19(1):10.
18. Alkaissi H, McFarlane SI. Artificial hallucinations in ChatGPT: implications in scientific writing. *Cureus.* 2023;15(2):e35074. doi:10.7759/cureus.35179
19. Thorp HH. ChatGPT is fun, but not an author. *Science*. 2023;379(6630):313. doi:10.1126/science.adg7879
20. Gravel J, D'Amours-Gravel M, Osmanlliu E. Learning to fake it: limited responses and fabricated references provided by ChatGPT for medical questions. *Mayo Clin Proc Digit Health*. 2023;1(3):226-234. doi:10.1016/j.mcpdig.2023.05.004
21. Orduña-Malea E, Cabezas-Clavijo Á. ChatGPT and the potential growing of ghost bibliographic references. *Scientometrics*. 2023;128(9):5351-5365. doi:10.1007/s11192-023-04804-4
22. Lechien JR, Briganti G, Vaira LA. Accuracy of ChatGPT-3.5 and -4 in providing scientific references in otolaryngology–head and neck surgery. *Eur Arch Otorhinolaryngol*. 2024;281(4):2159-2165. doi:10.1007/s00405-023-08441-8
23. Sanchez-Ramos L, Lin L, Romero R. Beware of references when using ChatGPT as a source of information to write scientific articles. *Am J Obstet Gynecol*. 2023;229(3):356. doi:10.1016/j.ajog.2023.04.004
24. Sebo P. How accurate are the references generated by ChatGPT in internal medicine? *Intern Emerg Med*. 2024;19(1):247-249. doi:10.1007/s11739-023-03484-5
25. Besançon L, Cabanac G, Labbé C, Magazinov A. Sneaked references: fabricated reference metadata distort citation counts. *J Assoc Inf Sci Technol*. 2024;75(12):1368-1379. doi:10.48550/arXiv.2310.02192
26. Walters WH, Wilder EI. Fabrication and errors in the bibliographic citations generated by ChatGPT. *Sci Rep*. 2023;13(1):14045. doi:10.1038/s41598-023-41032-5
27. Graf EM, McKinney JA, Dye AB, Lin L, Sanchez-Ramos L. Exploring the limits of artificial intelligence for referencing scientific articles. *Am J Perinatol.* 2024;41(15):2072-2081. doi:10.1055/s-0044-1786033
28. Altmäe S, Sola-Leyva A, Salumets A. Artificial intelligence in scientific writing: a friend or a foe? *Reprod Biomed Online*. 2023;47(1):3-9. doi:10.1016/j.rbmo.2023.04.009
29. Watson AP. Hallucinated citation analysis: delving into student-submitted AI-generated sources at the University of Mississippi. *Ser Libr.* 2024;85(5-6):172-180. doi:10.1080/0361526X.2024.2433640
30. Frosolini A, Franz L, Benedetti S, et al. Assessing the accuracy of ChatGPT references in head and neck and ENT disciplines. *Eur Arch Otorhinolaryngol.* 2023;280(11):5129-5133. doi:10.1007/s00405-023-08205-4
31. Giray L. ChatGPT references unveiled: distinguishing the reliable from the fake. *Internet Ref Serv Q.* 2024;28(1):9-18. doi:10.1080/10875301.2023.2265369
32. Athaluri SA, Manthena SV, Kesapragada VKM, Yarlagadda V, Dave T, Duddumpudi RTS. Exploring the boundaries of reality: investigating the phenomenon of artificial intelligence hallucination in scientific writing through ChatGPT references. *Cureus*. 2023;15(4):e37432. doi:10.7759/cureus.37432
33. Bhattacharyya M, Miller VM, Bhattacharyya D, Miller LE, Miller V. High rates of fabricated and inaccurate references in ChatGPT-generated medical content. *Cureus.* 2023;15(5):e39238. doi:10.7759/cureus.39238

Original Article

ANATOLIAN CURRENT MEDICAL

DOI: 10.38053/acmj.1753591

Anatolian Curr Med J. 2025;7(5):703-708

Educational quality of YouTube videos on enhancing lactation in postpartum mothers: a cross-sectional observational study

Mustafa Törehan Aslan[1], Mehmet Semih Demirtaş[2]

[1]*Department of Neonatology, Koç University Hospital, İstanbul, Turkiye*

[2]*Department of Pediatrics, Faculty of Medicine, Aksaray University, Aksaray, Turkiye*

Cite this article as: Aslan MT, Demirtaş MS. Educational quality of YouTube videos on enhancing lactation in postpartum mothers: a cross-sectional observational study. *Anatolian Curr Med J.* 2025;7(5):703-708.

Received: 29.07.2025 ◆ **Accepted**: 14.09.2025 ◆ **Published**: 15.09.2025

ABSTRACT

Aims: YouTube is a popular source of breastfeeding advice, but the quality of videos on increasing breast milk supply remains uncertain. We assessed the reliability and educational value of such content.

Methods: In this cross-sectional study, YouTube was searched between July 2024 and January 2025 for the keyword "increasing breast milk." An independent neonatologist and a social pediatrician evaluated each video using three validated tools: The DISCERN score (15-item quality measure, 1–5 per item; total 15–75), Journal of the American Medical Association (JAMA) benchmarks (0–4), and Global Quality Score (GQS, 1–5). Interrater reliability was evaluated with intraclass correlation.

Results: The 172 videos had a mean duration of 8.5±7.0 minutes (range 1.0–60.0 minutes) and a mean view count of 60.000±210.000 (8–1.700.000). Videos from academic institutions scored higher (DISCERN 45±12; JAMA 2.5±0.8; GQS 3.5±1.0) than those by individuals (DISCERN 25±8; JAMA 1.0±0.6; GQS 1.8±0.7; $p<0.001$). Educational videos had significantly greater quality than testimonial or promotional videos ($p<0.001$). Longer videos correlated strongly with quality ($r\approx0.55–0.60$, $p<0.001$), while views/likes showed weak associations ($r\approx0.20$, $p<0.05$). Comments had no correlation.

Conclusion: Most YouTube videos on breast milk supply are poor in quality and educational value. Trusted sources (academia, healthcare) provided better content, whereas individual/promotional videos were often unreliable. Healthcare providers should direct parents to credible resources and produce evidence-based materials to enhance online breastfeeding education.

Keywords: Breastfeeding, breastmilk, galactagog, newborn, YouTube, education

INTRODUCTION

Despite breastfeeding's well-established benefits, many mothers experience perceived insufficient milk supply-the leading cause of early cessation.[1,2] This concern drives mothers to seek lactation advice, increasingly through online platforms like YouTube. As the largest video-sharing platform, YouTube offers potential for accessible breastfeeding education through demonstrations and shared experiences, though the quality of such content warrants examination.[2,3]

The quality of health information on YouTube is highly inconsistent, with content often prioritized for popularity over accuracy.[4-6] Systematic reviews indicate that much of this material lacks evidence-based guidance, particularly in breastfeeding-related content. For instance, only 18.8% of YouTube breastfeeding videos were rated good/excellent, while most were suboptimal or misleading.[6] Similar issues plague pediatric health content, with studies reporting fewer than half of videos as accurate or comprehensive (e.g., only 56% accuracy in pediatric surgery videos). Such findings highlight the risks of unreliable or contraindicated advice for parents seeking lactation support online.[7,8]

To date, little academic attention has focused specifically on videos about increasing breast milk supply. While a recent study examined breastfeeding information on YouTube in Arabic,[3] and another analyzed content related to breastfeeding during the COVID-19 pandemic,[9] there remains a need to evaluate general content on enhancing milk production, particularly in English which has a global audience. Understanding the educational quality of these videos is important for healthcare providers who counsel postpartum mothers, as well as for identifying gaps where better resources are needed.

This study systematically evaluated the quality, reliability, and educational utility of YouTube videos on increasing breast milk supply using validated scoring tools. We analyzed how uploader type, content category, and engagement metrics influenced quality scores, hypothesizing that medical/academic sources would outperform individual uploaders and that overall quality would be variable but often suboptimal. By identifying strengths and gaps in current content, we aim to guide clinicians and educators on YouTube's utility as a patient resource and highlight areas for improvement.

Corresponding Author: Mustafa Törehan Aslan, torehanaslan@yahoo.com

METHODS

Ethics

We conducted a cross-sectional content analysis of YouTube videos. The study did not involve human subjects and was exempt from institutional review. All procedures were carried out in accordance with the ethical rules and the principles of the Declaration of Helsinki.

Video Selection

We searched YouTube between December 15, 2024 and February 10, 2025 for the term "increasing breast milk." Search results were sorted by relevance using a logged-out, non-personalized browser to minimize algorithmic bias. The first 200 video results were recorded. Videos were included if they were in English, focused on methods or tips for increasing or maintaining breast milk supply, and had a duration ≥60 seconds. We excluded videos that were shorter than 60 seconds (insufficient content), clearly unrelated to breastfeeding or lactation (off-topic), or had audio in an incomprehensible language or accent such that content could not be understood. After applying criteria, a total of 172 unique videos were included for analysis (**Figure 1**).

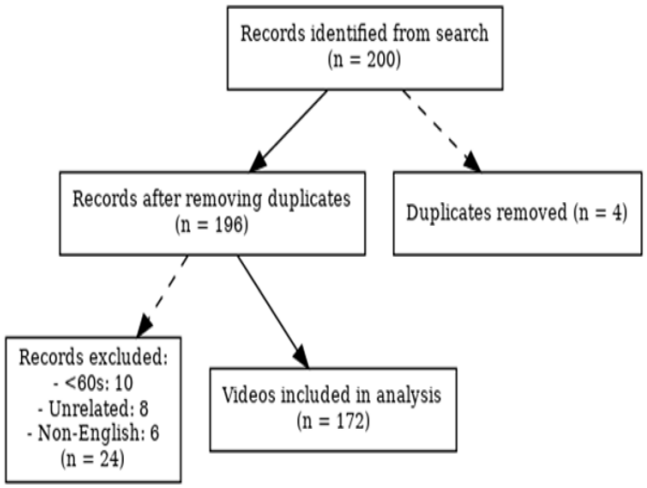


Figure 1. Flow chart of the study

Video Characteristics and Coding

For each included video, we recorded general metrics: upload date, video duration (seconds), view count, "likes" count, and number of comments. To account for varying ages of videos, we calculated engagement rates (views, likes, comments per day since upload) by dividing counts by the days since upload. We noted the uploader identity and categorized it into one of three uploader types: (1) Academic/Medical Institution (e.g., universities, hospitals, professional organizations), (2) Healthcare Professional (individual doctors, nurses, lactation consultants, or private clinics' channels), or (3) Individual (layperson or parent bloggers/vloggers, including non-professional influencers or commercial product promoters).

We also categorized each video's content type as educational (primarily instructional or factual information on increasing milk supply), testimonial (personal experiences or vlogs sharing one mother's journey or tips), or promotional (content primarily aimed at marketing a product or service, such as supplements or pumps, claimed to increase milk). Video continent of origin (based on uploader's stated location or video context) was recorded as North America, Europe, Asia, or other, to examine geographic trends. Two researchers independently performed the initial screening and categorization, with discrepancies resolved through discussion.

Quality Assessment Instruments

Three complementary instruments were used to evaluate video quality.

- **DISCERN:** A 15-question tool (score 15–75) rating health info quality (1–5 per item). Scores: <39 (poor), 39–50 (fair), >50 (good). Higher scores=better reliability.
- **JAMA Benchmarks:** Four criteria (authorship, attribution, disclosure, currency), each scored 0/1 (total 0–4). Higher scores=greater transparency.[10]
- **GQS:** A 5-point Likert scale rating video educational value (1=poor, 5=excellent). Measures understandability and usefulness for patients.[4,5,10]

A neonatologist and a social pediatrician with expertise in breastfeeding and lactation (blinded to each other's scores) independently viewed and scored all videos using the above instruments. Raters were first trained on 10 sample videos (not in the final dataset) to calibrate their scoring approaches. They were instructed to watch each video in full and could replay or pause as needed to apply the scoring criteria. No communication between raters occurred during the scoring phase. After scoring, interrater reliability was assessed by calculating the intraclass correlation coefficient for each scoring instrument's total scores across the two reviewers.

Statistical Analysis

Continuous variables were summarized with means, standard deviations (SD), and ranges, and categorical variables were summarized as frequencies and percentages. The Kolmogorov–Smirnov test was used to assess normality of score distributions; all primary variables met criteria for normal distribution ($p>0.05$), justifying use of parametric tests. We compared mean quality scores between two groups using independent-samples t-tests (e.g., academic vs non-academic uploader) and among three or more groups using one-way analysis of variance (ANOVA). For multi-group comparisons (uploader type, video type, region), post hoc pairwise comparisons were performed with the Bonferroni correction to identify which groups differed. Correlations between video numeric metrics (view count, like count, comment count, duration, and daily engagement rates) and quality scores were evaluated using Pearson's correlation coefficient (r). We interpreted correlation strength using standard guidelines: an r of 0.10–0.39 as weak, 0.40–0.69 as moderate, and ≥0.70 as strong correlation. Statistical significance was set at $p<0.05$ (two-tailed). All analyses were conducted using IBM SPSS Statistics (v28.0, IBM Corp).

RESULTS

Video Characteristics

The included videos were uploaded between 2015 and 2025, with a median upload year of 2020. On average, videos had been online for 4.1±3.3 years (range 50 days to 15.2 years). The mean total view count per video was 60.084±208.668 (range 8 to 1.700.000), with highly skewed distribution (median 5.421 views) reflecting a few extremely popular videos amid many with modest view counts. Videos received a mean of 500±1.500 likes (range 0 to 12.000) and 30±90 comments (range 0 to 500). Table 1 summarizes the descriptive characteristics of the videos. Regarding uploader type, 26 videos (15.1%) were from academic or medical institutions (e.g., university channels, professional associations, major hospitals).

Table 1. Characteristics of YouTube videos on increasing breast milk

Characteristic	Mean±SD/n (%)
Video duration (seconds)	508±420 (61–3600)
Days since upload	1490±1300 (50–5550)
Total views	60.084±208.668 (8–1.700.000)
Total likes	500±1500 (0–12.000)
Total comments	30±90 (0–500)
Daily views*	39.0±77.0 (0.1–511.0)
Daily likes*	0.13±0.22 (0–1.5)
Daily comments*	0.05±0.10 (0–0.8)
Source	**(n=172)**
Academic/medical institution	26 (15.1%)
Healthcare professional/clinic	69 (40.1%)
Individual (lay person)	77 (44.8%)
Video content	
Educational/informational	95 (55.2%)
Testimonial/personal story	52 (30.2%)
Promotional/commercial	25 (14.5%)
Broadcasted by region	
North America	80 (46.5%)
Asia	49 (28.5%)
Europe	35 (20.3%)
Other/Not clear	8 (4.7%)

SD: Standard deviation, *Daily values calculated as total count divided by days since upload (at time of analysis).

Interrater agreement between the neonatologist and social pediatrician reviewers was robust. The independent scores for each video were strongly correlated. The intraclass correlation coefficient was 0.994 for DISCERN, 0.879 for JAMA, and 0.898 for GQS (p<0.001 for all).

Quality and Reliability Scores (Overall)

Across all videos, the quality scores indicated generally poor reliability and educational content. The mean DISCERN score for the 172 videos was 34.2±12.5 (median 32; range 15–71 out of a maximum of 75). This falls in the category of "poor" quality consumer health information (well below the threshold of 39 for fair quality). In fact, 72% of videos scored <40 on DISCERN, highlighting that the majority lacked many of the elements of good-quality information (such as citing sources, discussing pros/cons of interventions, or areas of uncertainty). The mean JAMA benchmark score was 1.7±0.8 (median 2; range 0–3.5 out of 4). The mean GQS was 2.4±1.0 (median 2; range 0.5–5). Indeed, 60.5% of videos had GQS ≤2, reflecting that most were not very useful for patient education. Only 15 videos (8.7%) were rated as good or excellent (GQS ≥4).

Quality Scores by Uploader Category

There were significant differences in video quality depending on the uploader's identity (p<0.001 by ANOVA for DISCERN, JAMA, and GQS). Videos produced by academic institutions or professional medical organizations had the highest mean quality scores on all three instruments, followed by those from healthcare professionals, and lastly videos uploaded by individuals had the lowest scores (Table 2).

Table 2. Comparison of video quality scores by uploader type and video content type

Category	n	DISCERN mean±SD	JAMA mean±SD	GQS mean±SD
Academic/institution	26	45.1±11.5	2.5±0.8	3.5±0.9
Healthcare professional	69	35.3±9.8	1.8±0.7	2.5±0.8
Individual (lay person)	77	25.3±7.8	1.0±0.6	1.8±0.7
ANOVA p-value	-	<0.001	<0.001	<0.001
Educational	95	42.2±11.0	2.3±0.7	3.2±1.0
Testimonial	52	27.6±9.2	1.0±0.5	1.8±0.7
Promotional	25	29.6±10.2	1.2±0.6	2.0±0.8
ANOVA p-value	-	<0.001	0.002	<0.001

SD: Standard deviation, JAMA: Journal of the American Medical Association, GQS: Global Quality Score

Academic/institutional videos achieved a mean DISCERN of 45.1±11.5, which approaches "fair" quality (though still not consistently "good"). Post hoc comparisons showed that academic/institutional videos scored significantly higher than individual videos on all metrics (Bonferroni-adjusted p<0.001). Videos from healthcare professionals (mean DISCERN 35.3±9.8; JAMA 1.8±0.7; GQS 2.5±0.8) had intermediate quality. They were significantly better than individual videos (p<0.001 for DISCERN and GQS; p=0.014 for JAMA) but still notably lower than academic sources (Figure 2).

Quality Scores by Video Type

Video content classified as educational had substantially higher quality scores than testimonial or promotional videos (Table 2). Educational videos (generally those structured as lectures, tutorials, or Q&A with experts) had a mean DISCERN of 42.2±11.0, compared to 27.6±9.2 for testimonial videos and 29.6±10.2 for promotional videos (p<0.001 ANOVA). The pattern was similar for JAMA (mean 2.3 for educational vs 1.0 and 1.2 for testimonial and promotional, respectively, p=0.002) and GQS (3.2 vs ~1.9–2.0, p<0.001).

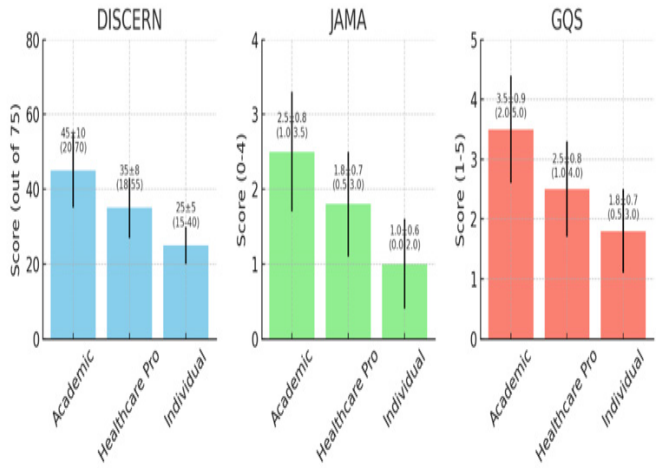


Figure 2. Mean quality scores (±SD) by uploader category
SD: Standard deviation, JAMA: Journal of the American Medical Association, GQS: Global Quality Score

Educational videos were significantly higher in quality than both testimonial and promotional videos on post hoc comparisons (all p<0.001) (**Figure 3**).

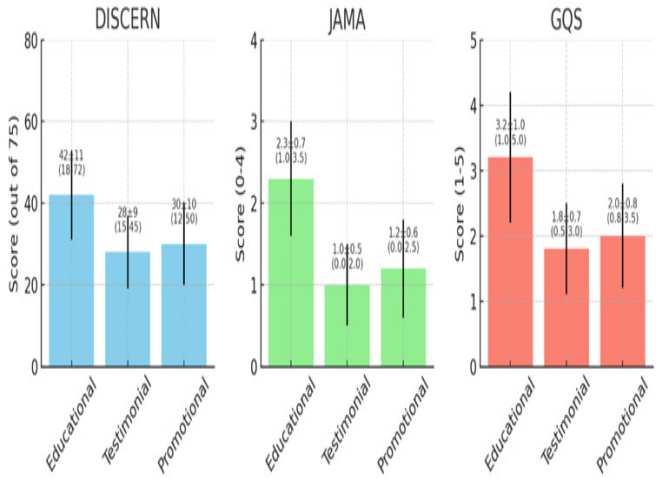


Figure 3. Mean quality scores (±SD) by video content type
SD: Standard deviation, JAMA: Journal of the American Medical Association, GQS: Global Quality Score

It was observed that the continent of video origin did not significantly affect video quality. mean DISCERN scores were similar for videos originating from North America, Europe, and Asia (ANOVA p=0.64), and no differences were found in JAMA or GQS by region (p=0.49 and p=0.72, respectively).

Viewer Engagement vs Quality

We examined whether videos with higher quality scores also showed greater viewer engagement (**Table 3**). There was a weak positive correlation between video quality and the number of likes a video received per day. Specifically, the Pearson correlation between daily likes and DISCERN was r=0.27 (p=0.001), indicating that higher-quality videos tended to garner more "likes" from viewers. One of the most striking findings was a strong positive correlation between video length (duration) and quality scores. Longer videos tended to have substantially higher DISCERN (r=0.60, p<0.001), JAMA (r=0.58, p<0.001), and GQS (r=0.55, p<0.001) scores. In fact, video duration showed the strongest association with quality out of all factors analyzed (**Table 3**).

Table 3. Correlation of viewer engagement metrics and video duration with quality scores

Parameter	r (DISCERN)	p (DISCERN)	r (JAMA)	p (JAMA)	r (GQS)	p (GQS)
Daily views	0.18	.021	0.09	.25	0.12	.15
Daily likes	0.27	.001	0.19	.014	0.22	.005
Daily comments	0.11	.16	0.08	.30	0.06	.45
Video duration (min.)	0.6	<.001	0.58	<.001	0.55	<.001

JAMA: Journal of the American Medical Association, GQS: Global Quality Score

DISCUSSION

In this cross-sectional analysis of YouTube content on increasing breast milk supply, we found that the overall quality of information is low, with relatively few videos meeting established criteria for reliability and completeness. The average DISCERN and JAMA scores of the 172 videos indicate poor-quality, patient-targeted information, echoing findings from prior studies of health information on YouTube.[6,11] This study is, to our knowledge, the first to specifically evaluate videos on augmenting breast milk production, and our results raise concern that mothers seeking help on this topic via YouTube may frequently encounter suboptimal advice.

Videos from academic/medical sources demonstrated superior quality (DISCERN 45.1±11.5 vs. 25±8 for individual uploaders, p<0.001), reflecting greater expertise in presenting balanced, evidence-based strategies (e.g., latch techniques, pumping frequency, nutritional guidance).[7,11,12] Such content frequently emphasized professional consultation, enhancing credibility. These findings align with prior research.[7,13,14] showing medical institutions produce more accurate content across specialties, though they constituted only 15% of our sample.

One positive observation is that videos from reputable sources–academic institutions and health professionals–scored significantly higher on quality metrics. This suggests that expertise matters: when experienced clinicians or organizations produce content, they are more likely to include balanced information, reference credible sources, and provide a more comprehensive overview of strategies. For instance, several high-scoring videos in our sample were from pediatric hospitals or well-known lactation consultants and covered multiple evidence-based techniques (ensuring proper latch, frequent nursing/pumping, treating maternal hydration and nutrition, managing stress and rest, possibly using galactagogues judiciously, etc.). These videos often mentioned the importance of consulting healthcare providers, which adds to their credibility. Our results mirror those of previous studies that found videos authored by professional organizations or medical authorities tend to be more accurate and complete.[11,15] For example, Bezner et al.[7] evaluated pediatric surgical videos and noted that those from medical institutions were far more likely to contain correct and comprehensive information than those from lay uploaders. Similarly, a very recent study of ear tube surgery videos found that content from universities and

hospitals had higher quality scores than that from private individuals as demonstrated in ophthalmologic YouTube content analyses.[12,14] Our study reinforces these patterns in the context of breastfeeding: expert-driven videos are comparatively better, though unfortunately they represent a minority of what is available (only ~15% of our sample).

Another key finding is the influence of video content type on quality. We saw that structured educational videos vastly outperformed testimonials and promotional content in quality metrics. Educational videos often took a didactic approach, sometimes with slides or a talking-head expert explaining various aspects of lactation management. These videos were more likely to mention the biology of milk production, strategies like skin-to-skin contact, frequency of feeding/pumping, checking infant latch, and when to seek professional help (e.g., for possible tongue-tie or other issues). In contrast, testimonial videos, while engaging, usually focused on the individual's narrative and often omitted broader guidance–reflected in very low DISCERN scores (many scored ≤25). Promotional videos had slightly better scores than testimonials in our sample (though differences were not statistically significant between those two groups), possibly because some promotional content included semi-formal presentations about their product's benefits. Nevertheless, promotional videos inherently present biased information; they rarely acknowledged alternative solutions a concern echoed in other surgical domains such as podiatry.[16] or any drawbacks of the promoted intervention, leading to low DISCERN sub-scores for balance and discussion of alternatives. The net effect is that a mother watching mainly testimonials or ads would receive a limited scope of suggestions, potentially skewed by personal bias or commercial interest. This underscores a concern that has been raised across various medical fields on YouTube: content intended to market or persuade often compromises on completeness and impartiality.[11]

It is worth noting that video length showed a strong correlation with quality a pattern also observed in oculoplastic educational content.[17] Longer videos tended to be more comprehensive and thus scored higher on instruments that reward completeness of information. This suggests a practical insight: a video that is just 2–3 minutes long is unlikely to adequately cover the multifaceted topic of increasing breast milk supply. Indeed, many very short videos in our dataset provided only superficial advice (e.g., "drink more water and try fenugreek") without elaboration. On the other hand, videos in the ~15-20 minute range often included segments on multiple tips (like proper latch technique demonstration, pumping strategies, dietary advice, etc.), aligning with more criteria on the DISCERN checklist. This correlation should not be misinterpreted as length causing quality, but rather as an indicator that to present quality information one needs to devote sufficient time. Prior analyses in other domains have similarly found that videos addressing a medical topic in depth (which naturally makes them longer) tend to achieve higher reliability scores.[11] However, longer videos may also demand more viewer commitment and might not be fully watched by all users, which raises a challenge: the highest-quality content is not always in a viewer-friendly format for quick consumption. This tension between depth and viewer engagement might partly explain why we did not find strong correlations between view counts (or likes) and quality–viral popularity favors brevity and catchiness over thoroughness, whereas quality favors thoroughness over brevity.[6]

Our study has important implications for clinical practice and patient education. First, healthcare providers should be aware that patients may form impressions or follow advice from YouTube videos that are not evidence-based. Many postpartum mothers struggling with milk supply turn to platforms like YouTube out of convenience or desperation. If a mother reports she has been trying certain remedies she saw online (such as herbal supplements, lactation cookies, or extreme pumping regimens), clinicians should gently inquire about the source and help her distinguish which advice is sound and which may be unproven or counterproductive. For example, excessive emphasis on lactation teas could distract from more effective measures like improving latch or frequency of feeds. Pediatricians, neonatologists, and lactation consultants may need to proactively guide mothers to reputable online resources. There are a few high-quality videos (some from this study's sample) produced by certified lactation consultants and medical centers – sharing links to those, or to trusted organizations (La Leche League, UNICEF breastfeeding videos, etc.), could help patients get better information. Additionally, our findings suggest that professional organizations could consider creating more engaging, shareable content to fill the current void. Given that expert videos were relatively few, increasing their presence on YouTube might improve the overall quality mix available to the public.

For the YouTube platform and content creators, our results reinforce the recommendation that popularity should not be the sole metric of usefulness. The weak correlation between view counts/likes and quality underscores that viewer engagement metrics do not reliably signal accuracy.[6] This has been noted in prior research and was echoed by the systematic review authors who suggested incorporating expert evaluations into video rankings.[6] YouTube's algorithms could potentially be adapted to elevate content that meets certain quality criteria (perhaps via partnerships where health institutions are verified). In the meantime, creators of breastfeeding content who want to provide value should consider collaborating with healthcare professionals to ensure accuracy, and include references or citations for claims (which was rarely done in our sample, but would improve JAMA scores and trust). Even simple steps like stating one's credentials and citing sources (e.g., World Health Organisations (WHO) recommendations,[18] for breastfeeding could improve a video's credibility.

Finally, it is notable that our analysis did not find regional differences in quality–the misinformation problem transcends borders. Whether a video was made in the U.S. or in India, its quality depended on the content and creator rather than the country. This suggests that global collaboration and standards may be beneficial. International lactation consultant associations or pediatric societies could work together to

produce multi-language video content that meets high quality standards, to serve as reliable alternatives across regions.

Limitations

This study has important limitations. The sample size, though focused on top-viewed content, was limited and a broader analysis would strengthen the findings. As a snapshot in time, this study cannot capture the constantly changing nature of YouTube. Our exclusion of non-English videos limits the cross-cultural relevance of the results. While we used validated tools, assessing video quality involves some subjectivity. Crucially, we did not verify the medical accuracy of the advice or measure the videos' real-world impact on breastfeeding behaviors.

CONCLUSION

As a result, our evaluation shows that while YouTube contains some high-quality videos on increasing breast milk supply, the majority are neither reliable nor comprehensive. Videos from academic and healthcare-affiliated sources offer the best information but are vastly outnumbered by lower-quality personal or promotional videos. Many mothers seeking help online may therefore be at risk of receiving incomplete or misleading advice. It is crucial for healthcare professionals to be cognizant of the information their patients may encounter and to help direct them to trustworthy resources. Efforts should be made to increase the presence and visibility of accurate, evidence-based breastfeeding educational videos on platforms like YouTube. Multidisciplinary collaboration between medical experts, lactation consultants, and skilled communicators could yield content that is both engaging and informative, to better support breastfeeding mothers in the digital age. As online media becomes increasingly influential inpatient health behaviors, ensuring the quality of such information is a responsibility that the medical community cannot afford to overlook.

ETHICAL DECLARATIONS

Ethics Committee Approval

The study did not involve human participation or animal testing, and since the information was obtained from a publicly available platform, no ethics committee approval was required.

Informed Consent

The study did not involve human participation or animal experiments, and informed consent was not required as the information was obtained from a publicly available platform.

Referee Evaluation Process

Externally peer-reviewed.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Financial Disclosure

The authors declared that this study has received no financial support.

Author Contributions

All of the authors declare that they have all participated in the design, execution, and analysis of the paper, and that they have approved the final version.

REFERENCES

1. Demirtas MS, Yalcin SS. The use of human milk for therapeutic purposes other than nutrition. *Turk Arch Pediatr.* 2022;57(3):255-266. doi:10.5152/TurkArchPediatr.2022.22075
2. Piccolo O, Kinshella MW, Salimu S, et al. Healthcare worker perspectives on mother's insufficient milk supply in Malawi. *Int Breastfeed J.* 2022; 17(1):14. doi:10.1186/s13006-022-00460-1
3. Almoayad F, Alhashem A, Alotaibi R, et al. YouTube video clips on breastfeeding education and promotion for Arabic-speaking populations: a social media content analysis. *Breastfeed Med.* 2024;19(9): 734-741. doi:10.1089/bfm.2024.0149
4. Demirtas MS, Alici N. YouTube as a source of information on infantile colic. *Pediatr Int*. 2023;65(1):e15624. doi:10.1111/ped.15624
5. Demirtas MS, Alici N. The reliability and quality of YouTube videos as a source of breath holding spell. *Ital J Pediatr.* 2024;50(1):8. doi:10.1186/s 13052-023-01570-0
6. Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. *BMC Med Educ*. 2022;22(1):382. doi:10.1186/s12909-022-03446-z
7. Bezner SK, Hodgman EI, Diesen DL, et al. Pediatric surgery on YouTube: is the truth out there? *J Pediatr Surg.* 2014;49(4):586-589. doi:10.1016/j. jpedsurg.2013.08.004
8. Sehgal M, Jain D, Goel P, et al. An exploratory study on pediatric inguinal hernia videos on the YouTube platform. *J Indian Assoc Pediatr Surg.* 2022;27(5):588-593. doi:10.4103/jiaps.jiaps_240_21
9. Azak M, Yilmaz B, Sahin N. Analysis of YouTube((c)) videos regarding breastfeeding during the coronavirus disease (COVID-19) pandemic. *Matern Child Health J.* 2023;27(9):1548-1558. doi:10.1007/s10995-023-03722-4
10. Demirtas MS, Kocagozoglu SG, Yalcin SS. YouTube as a source of information about air pollution on children's health: importance for public health. *Int J Environ Health Res*. 2025:pp.1-10. doi:10.1080/09603123. 2025.2474089
11. Alacamli G, Yakar K. The reliability of YouTube as an information resource for parents about retinopathy of prematurity. *J Pediatr Ophthalmol Strabismus.* 2025;62(1):42-49. doi:10.3928/01913913-20240911-04
12. Aktar Ugurlu G, Ugurlu BN. Evaluating the quality and reliability of youtube videos on tympanostomy tubes: a comprehensive analysis for patients and parents. *BMC Public Health*. 2025;25(1):776. doi:10.1186/s 12889-025-21963-6
13. Knight K, van Leeuwen DM, Roland D, Moll HA, Oostenbrink R. YouTube: are parent-uploaded videos of their unwell children a useful source of medical information for other parents? *Arch Dis Child*. 2017; 102(10):910-914. doi:10.1136/archdischild-2016-311967
14. Uyar E, Saribas F. The assessment of YouTube videos as a source of information for amblyopia treatment. *J Pediatr Ophthalmol Strabismus.* 2021;58(5):311-318. doi:10.3928/01913913-20210409-01
15. Orbatu D, Yildirim Karaca S, Alaygut D, Karaca I. Educational features of YouTube videos depicting breastfeeding: quality, utility, and reliability analysis. *Breastfeed Med.* 2021;16(8):635-639. doi:10.1089/ bfm.2020.0301
16. Sari E, Umur LF. Quality analysis of hallux valgus videos on YouTube. *J Am Podiatr Med Assoc.* 2021;111(5):10.7547/20-191. doi:10.7547/20-191
17. Garip R, Sakallioglu AK. Evaluation of the educational quality and reliability of YouTube videos addressing eyelid ptosis surgery. *Orbit.* 2022;41(5):598-604. doi:10.1080/01676830.2021.1989467
18. Exclusive breastfeeding rates by country (Global Breastfeeding Scorecard 2023). WHO; 2024.