T.C. SAĞLIK BAKANLIĞI HASEKİ EĞİTİM VE ARAŞTIRMA HASTANESİ

The Medical Bulletin of Haseki

E-ISSN: 2147-2688

2026
Volume 64
Issue
March 2

www.hasekidergisi.com

The Medical Bulletin of Haseki

Editorial Board

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

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

The Medical Bulletin of Haseki

Scientific Advisory Board

The Medical Bulletin of Haseki

Scientific Advisory Board

The Medical Bulletin of Haseki

Scientific Advisory Board

The Medical Bulletin of Haseki

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

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

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

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

Responsible Manager: Akif ERBİN

The Medical Bulletin of Haseki

Contents

Original Articles

Original Article

DOI: 10.4274/haseki.galenos.2026.58561
Med Bull Haseki 2026;64(2):76-84

Deep Learning-based Differentiation of Idiopathic Granulomatous Mastitis from Malignant Non-mass Enhancement Using Breast Magnetic Resonance Imaging

Filiz Tasci[1], Esat Kaba[1], Mahmut Nedim Ekersular[2], Ahmet Alkan[2], Huseyin Er[3], Nur Hursoy[1]

[1]*Recep Tayyip Erdogan University Training and Research Hospital, Department of Radiology, Rize, Türkiye*

[2]*Kahramanmaras Sutcu Imam University Faculty of Engineering and Architecture, Department of Electrical and Electronics Engineering, Kahramanmaras, Türkiye*

[3]*Erzincan Binali Yildirim University Faculty of Medicine, Department of Radiology, Erzincan, Türkiye*

Abstract

Aim: Idiopathic granulomatous mastitis (IGM) is a benign, chronic inflammatory disease of the breast, and its imaging findings may overlap with those of malignant non-mass enhancement (NME). This study aimed to investigate the performance of deep learning and machine learning models in differentiating IGM from malignant NME based on dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI).

Methods: In this retrospective study conducted between January 2019 and March 2023, DCE-MRI findings of 30 patients with histopathologically confirmed IGM and of 33 patients with breast cancer presenting as NME were analyzed. The second dynamic phase of DCE-MRI (Dataset 1, 475 images) and the corresponding subtracted images (Dataset 2, 402 images) were used in this study. Datasets were sequentially split into 80% training and 20% testing sets to ensure a patient-level split. Image features were extracted using SqueezeNet and classified with a narrow neural network.

Results: The mean age was significantly lower in the IGM group than in the NME group (41.3±11.3 vs. 52.2±11.4 years, $p<0.001$). For Dataset 1, the area under the curve was 0.997 in training and 0.870 in testing; for Dataset 2, the area under the curve was 0.998 in training and 0.807 in testing. Training accuracy was 0.984 (Dataset 1) and 0.978 (Dataset 2), whereas test accuracy was 0.811 (Dataset 1) and 0.704 (Dataset 2).

Conclusion: The findings of this study suggest that deep learning shows significant promise for non-invasive differentiation of IGM from malignant NME on DCE-MRI, particularly in cases that are clinically indistinguishable.

Keywords: Granulomatous mastitis, breast neoplasms, magnetic resonance imaging, deep learning

Introduction

Idiopathic granulomatous mastitis (IGM), also known as granulomatous lobular mastitis, is a rare, benign, recurrent, and chronic inflammatory disease of the breast. It predominantly affects premenopausal women with a history of pregnancy and lactation (1,2). The disease is characterized by the formation of non-necrotizing granulomas involving the breast lobules, often accompanied by microabscesses, without evidence of microbial infection. Despite being a benign condition, IGM represents a significant diagnostic challenge because it can clinically and radiologically mimic infectious mastitis,

Corresponding Author: Esat Kaba, MD, Recep Tayyip Erdogan University Training and Research Hospital, Department of Radiology, Rize, Türkiye
E-mail: esatkaba04@gmail.com **ORCID:** orcid.org/0000-0001-7464-988X
Received: 09.02.2026 **Accepted:** 04.03.2026 **Publication Date:** 27.03.2026

Cite this article as: Tasci F, Kaba E, Ekersular MN, Alkan A, Er H, Hursoy N. Deep learning-based differentiation of idiopathic granulomatous mastitis from malignant non-mass enhancement using breast magnetic resonance imaging. Med Bull Haseki. 2026;64(2):76-84

inflammatory breast carcinoma, and other granulomatous diseases (3,4). Therefore, accurate diagnosis often requires a multidisciplinary approach involving radiologists, pathologists, and clinicians to avoid diagnostic delays and unnecessary interventions. A comprehensive evaluation typically includes imaging, clinical correlation, and histopathological examination, the latter remaining the gold standard for confirming IGM (5,6).

Magnetic resonance imaging (MRI), particularly dynamic contrast-enhanced MRI (DCE-MRI), plays a pivotal role in determining the extent of inflammation, detecting associated complications such as abscesses or fistulas, and differentiating IGM from other breast pathologies by providing detailed soft-tissue characterization and vascular enhancement patterns (7). Dynamic contrast-enhanced MRI enables evaluation of lesion distribution, enhancement characteristics, and kinetic curves, which are critical for breast lesion assessment. However, DCE-MRI findings in IGM may overlap with those of invasive malignancies and present as non-mass enhancement (NME), closely mimicking malignant NME patterns. This overlap can significantly complicate differential diagnosis, leading to misinterpretations and unnecessary biopsies (7,8).

We hypothesized that applying deep learning and machine learning models to breast MRI could effectively differentiate IGM from NME, thereby improving diagnostic accuracy in challenging cases. Accordingly, this study aimed to evaluate the diagnostic performance of these models in differentiating IGM from malignant NME. Thus, this approach may reduce diagnostic uncertainty, support clinical decision-making, and potentially prevent unnecessary invasive procedures.

Materials and Methods

Compliance with Ethical Standards

The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki, and approval was obtained from the Non-Interventional Clinical Research Ethics Committee of Recep Tayyip Erdogan University Faculty of Medicine (approval no: 2023/107, date: 30.03.2023). Written informed consent was obtained from all patients before undergoing breast DCE-MRI.

Study Design and Datasets

This retrospective study was conducted at a tertiary radiology clinic between January 2019 and March 2023 and included 30 female patients (mean age: 41.30 years, range 25 to 69 years) with ultrasonography and DCE-MRI findings suggestive of IGM, which were pathologically confirmed as IGM, and 33 female patients (mean age: 52.21 years, range 34 to 82 years) with DCE-MRI findings suggestive of malignant NME, pathologically confirmed as breast cancer. The exclusion criteria were poor image quality, suboptimal contrast timing, and unconfirmed pathological diagnoses. After applying these criteria, a total of 63 patients were included in the study. Two datasets were designed using MRI images from these patients: Dataset 1 comprised T1-weighted (T1W) contrast-enhanced images from the second phase of the patients' dynamic breast MRI, classified into two categories. This dataset included 30 patients in the IGM class and 33 in the NME class, for a total of 475 images (IGM=237, NME=238). Dataset 2 was created from subtracted images obtained from the same dynamic series of patients as in dataset 1, maintaining the same classifications, and included a total of 402 images (IGM=194, NME=208). The datasets utilized in this study exhibited a relatively equitable distribution between the IGM and NME classes, mitigating the potential influence of class imbalance on model training and assessment. The diagnoses of all the patients included in the study were pathologically confirmed.

Magnetic Resonance Imaging Parameters

Magnetic resonance imaging was performed using a 3.0-T magnetic resonance device (GE Healthcare Discovery MR750, Waukesha, WI, USA) with a 16-channel dedicated breast coil. The patients were positioned prone, with the breasts placed within the breast coil. A survey sequence was followed, for both breasts, by an axial T1W sequence and a fat-saturated T2-weighted (T2W) fast spin-echo sequence before contrast administration to avoid signal alteration due to the injected gadolinium. For DCE-MRI, the contrast agent gadobutrol (Gadovist, Bayer Schering Pharma, Berlin, Germany) was injected as a 0.1 mmol/kg bolus at a flow rate of 2 mL/s. After the injection, six phases of volume imaging for breast evaluation (VIBRANT-Flex) were employed, with approximately 60-s intervals between each phase and a total scanning of 410 s (repetition time, 3.9 ms; shortest echo time; flip angle, 12; field of view, 360-360 mm; matrix, 320-320; and layer thickness, 1.4 mm).

Classification

Datasets 1 and 2 were each divided into two groups, with the first 80% of the images in each dataset reserved for classifier training and validation and the remaining 20% allocated to classifier testing. Rather than randomly selecting images for an 80-20% split, images were selected from the beginning, allowing the classifier to evaluate patient slices it had not seen before during the testing phase. The datasets were sequentially split into 80% training and 20% testing sets to ensure a patient-level split. To prevent possible data leakage from multiple slices from the same patient, the dataset was divided at the patient level instead of the image level. All of a patient's images were put into either the training set or the test set so that no images from the same patient were in both sets.

Within the training set, five-fold cross-validation was performed to optimize model performance and reduce the risk of overfitting. The dataset was partitioned into five mutually exclusive subsets. The model was trained on four subsets and validated on the remaining subset; this process was repeated five times so that each subset served once as the validation set (9). The independent test set was not involved in the cross-validation process.

A hybrid framework combining deep learning–based feature extraction and machine learning–based classification was adopted. For each image, 1,000 deep features were extracted using the pre-trained convolutional neural network SqueezeNet, which was originally trained on over one million images from the ImageNet dataset and consists of 68 layers. The network accepts input images of 227×227 resolution and extracts from the pool10 layer, yielding a 1×1×1000 feature vector (10).

In the classification phase of the feature vectors extracted by the deep convolutional neural network (SqueezeNet), a structured narrow neural network architecture was preferred. Architecturally, this network is a feedforward multilayer perceptron with a single hidden layer. The architectural features and parameters of the network were configured as follows:

Input layer: It is designed to directly accept the 1*1000-dimensional feature vector extracted from the pool10 layer of SqueezeNet, which represents a patient's image.

Hidden layer: To keep the computational complexity of the model low and to prevent overfitting of high-level features extracted from the already-deep network (SqueezeNet), only 10 neurons were used in the hidden layer. The Rectified Linear Unit [ReLU, f(x)=max(0, x)], which prevents the vanishing gradient problem, was preferred as the activation function for the neurons in this layer.

Output layer: Since a binary classification problem (IGM and NME) was addressed, the output layer consists of 2 neurons. In this layer, the Softmax activation function was used to calculate the probability distribution of the features belonging to each class. During model training, the cross-entropy loss function was minimized to reduce error.

This narrow and shallow network structure offers superior performance in modeling non-linear relationships in high-dimensional data obtained from deep networks compared with traditional classifiers such as support vector machines and k-nearest neighbors, while eliminating the high hardware and time costs required by multi-layer deep networks (11). The overall workflow of the study is summarized in the flowchart presented in Figure 1. In addition, Figure 2 illustrates representative DCE-MRI findings from patients in the dataset.

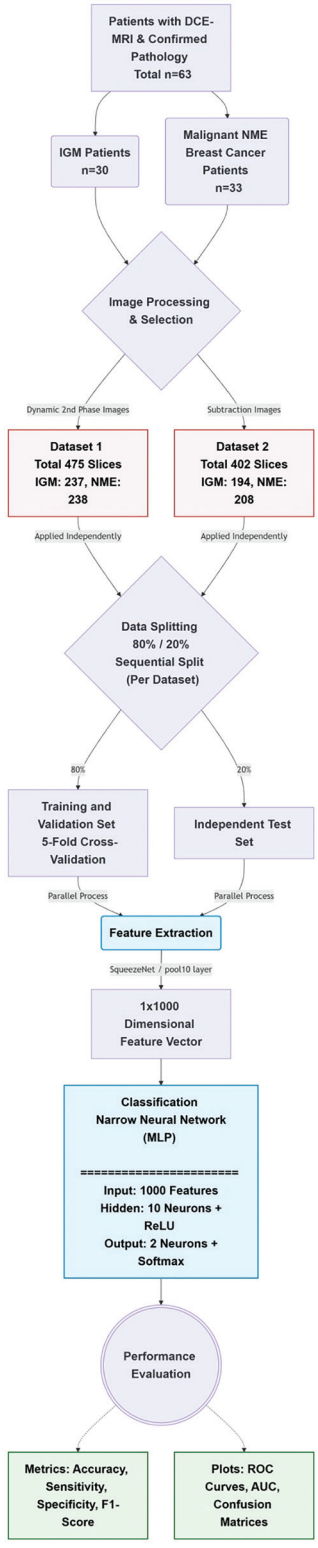

Figure 1. Flowchart of the study
IGM: Idiopathic granulomatous mastitis, NME: Non-mass enhancement, DCE-MRI: Dynamic contrast-enhanced magnetic resonance imaging, ReLU: Rectified Linear Unit, ROC: Receiver operating characteristic, AUC: Area under the curve

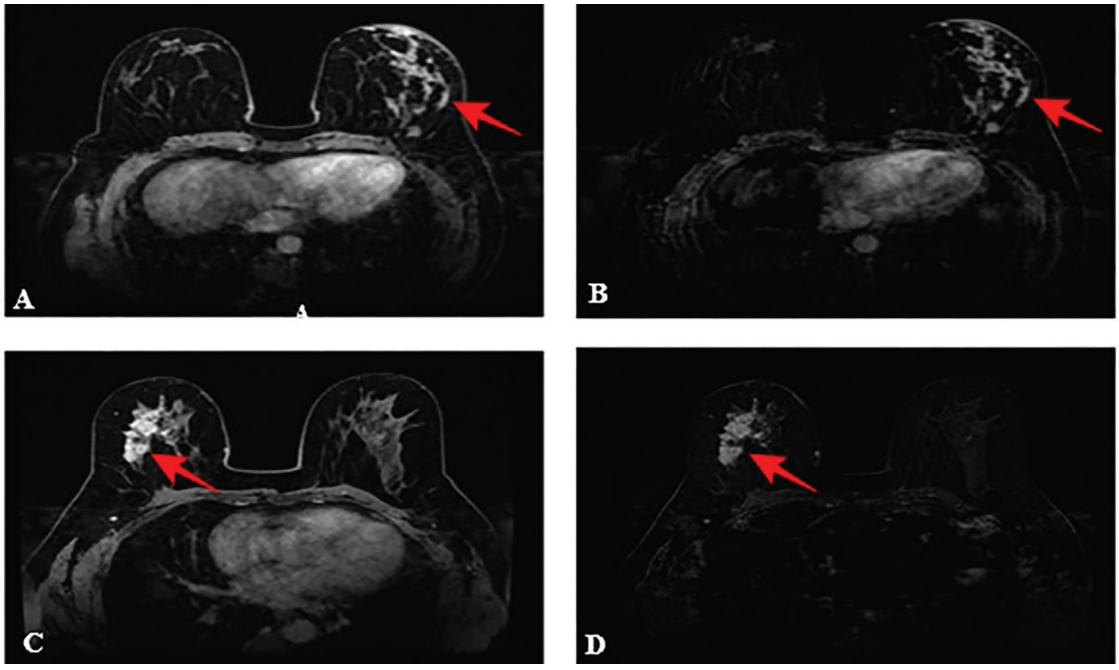

Figure 2. Magnetic resonance images. The fat-suppressed post-contrast T1-weighted image (**A**), and subtraction image (**B**) of a 45-year-old patient diagnosed with IGM show non-mass enhancement (red arrows) without clear borders in the retroareolar and outer quadrant of the left breast. The fat-suppressed post-contrast T1-weighted image (**C**) and subtraction image (**D**) show a lesion in the retroareolar and outer quadrants of the right breast in a 57-year-old patient diagnosed with invasive lobular carcinoma (red arrows)
IGM: Idiopathic granulomatous mastitis

Statistical Analysis

Statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS Inc., Chicago, IL) for Windows, version 23.0, and Python libraries, including NumPy, Pandas, Scikit-learn, and Matplotlib. The difference in age between the IGM and NME groups was evaluated using the Mann-Whitney U test. Age differences among NME pathological subtypes were analyzed using the Kruskal-Wallis test. Descriptive statistics were reported as mean ± standard deviation or median (minimum-maximum), as appropriate. Statistical significance was defined as a two-tailed p-value <0.05. The model's performance metrics were calculated separately for the validation and test datasets. Sensitivity, specificity, accuracy, precision, and F1-score were calculated from the components of the confusion matrix. Ninety-five percent confidence intervals (95%) for performance metrics were calculated using the Wilson score method. Receiver operating characteristic (ROC) curves and area under the curve (AUC) values were generated to assess discriminative performance.

Results

The mean age of the 30 patients with IGM was 41.30±11.31 (age range: 25-69) years, and that of the 33 patients with NME was 52.21±11.43 (age range: 34-82) years (p<0.001) (Table 1). The distribution of the pathology diagnoses of the patients with breast cancer in the NME class was as follows: invasive lobular carcinoma (n=12, mean age: 56.42±12.56 years), invasive ductal carcinoma (n=9, mean age: 52.78±13.48 years), invasive micropapillary carcinoma (n=7, mean age: 50.14±5.55 years), and ductal carcinoma *in situ* (n=5, mean age: 44.0±7.07 years) (p=0.198) (Table 2).

Confusion matrices and ROC graphs of the narrow neural network classifier were obtained to evaluate the classification performance on Datasets 1 and 2. These evaluations were also undertaken separately for the validation and test images. The confusion matrices for the images in Dataset 1 are shown in Figure 3. The confusion matrix illustrating the classification performance on the validation data (Figure 3a) indicates that the narrow neural network classifier correctly classified 186 of the 190 IGM images; four were misclassified as NME. The same matrix shows that the classifier correctly classified 188 of the 190 NME images as NME, with the remaining two misclassified as IGM. Examination of the confusion matrix for the test set (Figure 3b) of the same dataset shows that 38 of the 47 IGM images were correctly classified as IGM, and nine of the 48 NME images were incorrectly classified as IGM.

Figure 4 presents the classification values for the images included in Dataset 2. The confusion matrix for the validation images (Figure 4a) shows that the classifier correctly identified 152 of 155 IGM images (three misclassified as NME) and 162 of 166 NME images (four

misclassified as IGM). The confusion matrix for the test set of Dataset 2 (Figure 4b) shows that 34 of the 42 NME images were correctly classified as NME, and eight were misclassified as IGM.

Figure 5 shows the ROC curves and the AUC values for the validation phase of Datasets 1 and 2, where the positive class is defined as IGM. On the ROC curve, the true positive rate is plotted on the y-axis, and the false positive rate is plotted on the x-axis to illustrate classifier performance. The AUC indicates the classifier's ability to distinguish between classes. An AUC value approaching 1 indicates that the classifier is highly successful in differentiating between classes. For Datasets 1 and 2, the validation phase yielded AUC values of 0.997 and 0.998, respectively. These high AUC values demonstrate that the narrow neural network classifier successfully predicted class membership.

Figure 6 presents the ROC plots and AUC values obtained from the test images. For the classification of the images included in Datasets 1 and 2, which were allocated for testing, the AUC values were 0.870 and 0.807, respectively.

Other measurement values that show the classification performance of the narrow neural network classifier during the validation phase for both datasets are given in Table 3. For this phase, the accuracy values of Datasets 1 and 2 were 0.984 and 0.978, respectively.

Table 4 shows the classification performance metrics for the two datasets derived from images allocated for testing. The accuracy values achieved during the test phase were 0.811 for Dataset 1 and 0.704 for Dataset 2.

Table 1. Mean ages of the NME and IGM groups

Age	NME	IGM	p*
Mean±SD	52.21±11.43	41.30±11.31	<0.001
Median (min.-max.)	50 (34-82)	39 (25-69)	

*Mann-Whitney U test
NME: Non-mass enhancement, IGM: Idiopathic granulomatous mastitis, SD: Standard deviation, min.: Minimum, max.: Maximum

Table 2. Mean ages of the NME subtypes

	ILC (n=12)	IDC (n=9)	IMC (n=7)	DCIS (n=5)	p*
Age					
Mean±SD	56.42±12.56	52.78±13.48	50.14±5.55	44.0±7.07	0.198
Median (min.-max.)	51 (42-81)	51 (47-82)	47 (44-56)	45 (34-53)	

*Kruskal-Wallis test
NME: Non-mass enhancement, ILC: Invasive lobular carcinoma, IDC: Invasive ductal carcinoma, IMC: Invasive micropapillary carcinoma, DCIS: Ductal carcinoma *in situ*, SD: Standard deviation, min.: Minimum, max.: Maximum

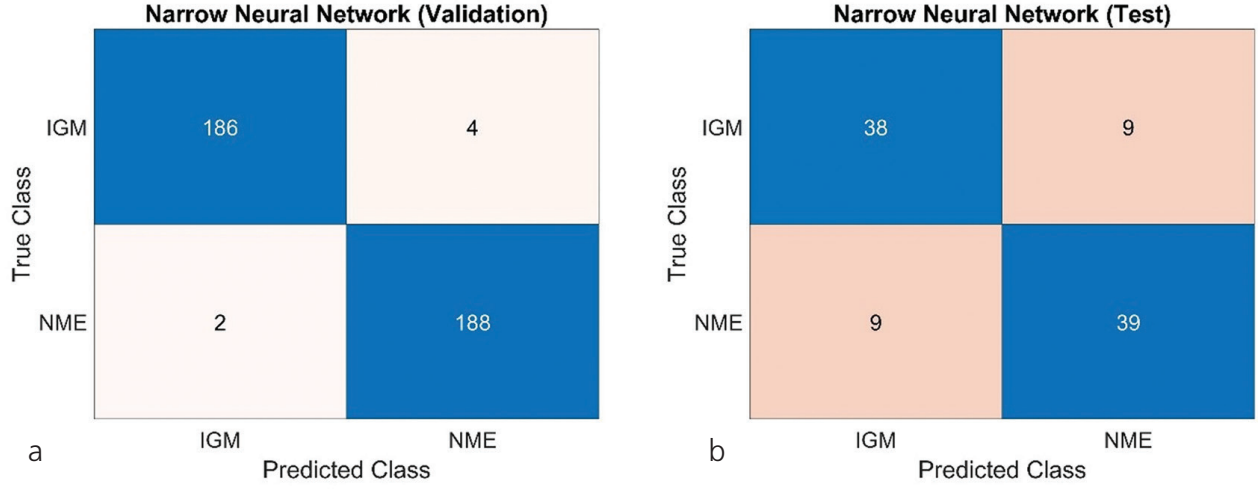

Figure 3. Confusion matrices of the validation (**a**) and test (**b**) data for Dataset 1
IGM: Idiopathic granulomatous mastitis, NME: Non-mass enhancement

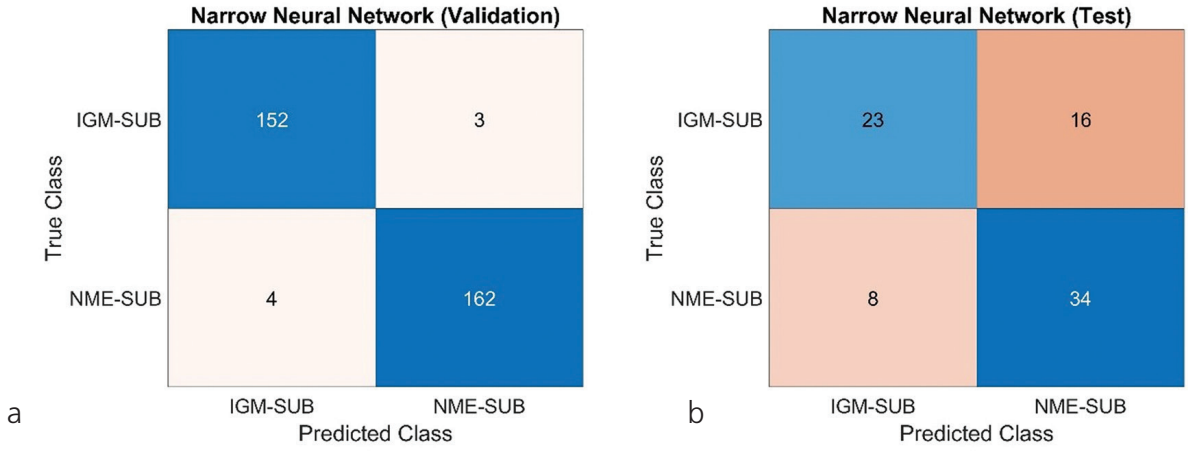

Figure 4. Confusion matrices of the validation (**a**) and test (**b**) data for Dataset 2
IGM: Idiopathic granulomatous mastitis, NME: Non-mass enhancement

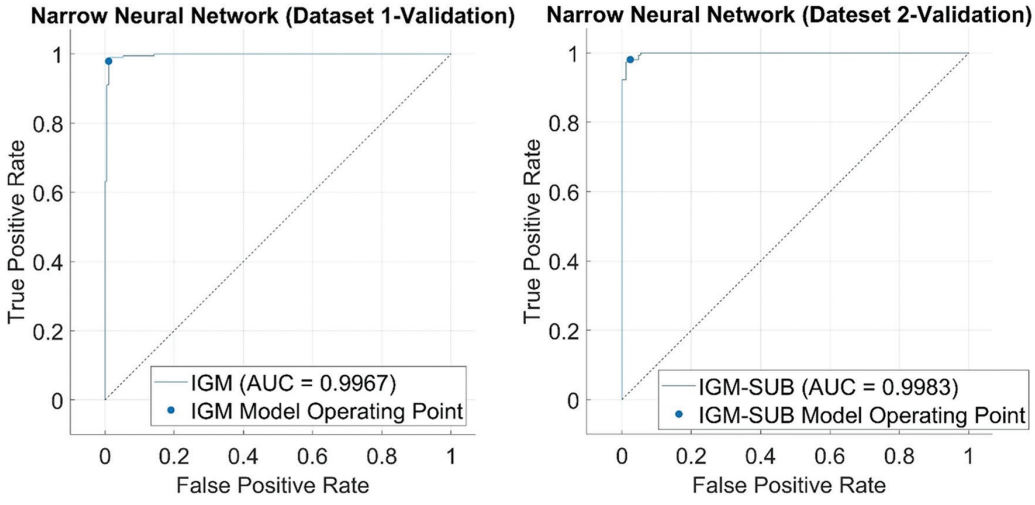

Figure 5. Receiver operating characteristic graphs and area under the curve values of the validation phase (for Datasets 1 and 2)
IGM: Idiopathic granulomatous mastitis, AUC: Area under the curve

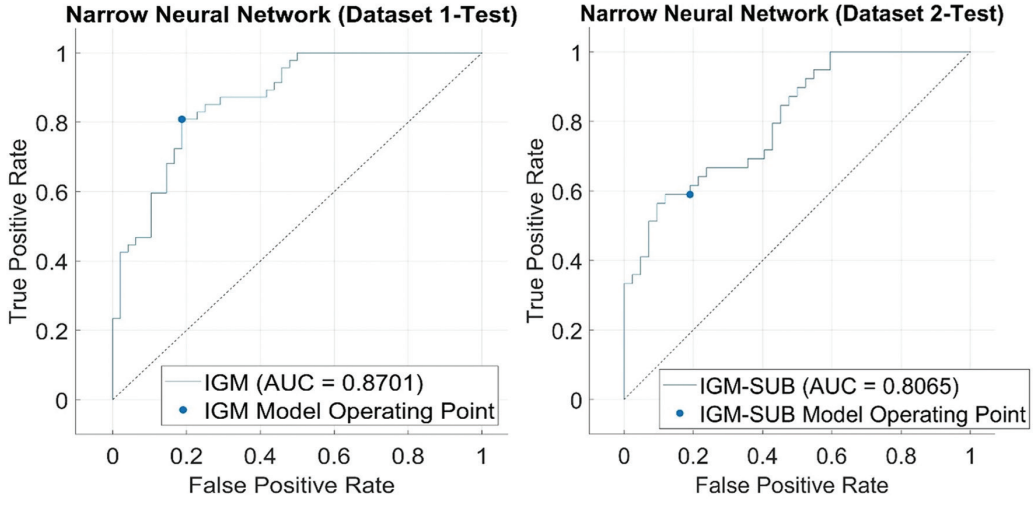

Figure 6. Receiver operating characteristic graphs and area under the curve values of the test phase (for datasets 1 and 2)
IGM: Idiopathic granulomatous mastitis, AUC: Area under the curve

Table 3. Classification performance metrics of the narrow neural network classifier for the validation images					
	Accuracy (95% CI)	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Precision (95% CI)**	**F1-score**
Dataset 1 (validation)	0.984 (0.965-0.993)	0.979 (0.946-0.992)	0.990 (0.959-0.998)	0.989 (0.959-0.998)	0.984
Dataset 2 (validation)	0.978 (0.954-0.989)	0.981 (0.944-0.995)	0.976 (0.938-0.991)	0.974 (0.936-0.990)	0.978
CI: Confidence interval					

Table 4. Classification performance metrics of the narrow neural network classifier for the test images					
	Accuracy (95% CI)	**Sensitivity (95% CI)**	**Specificity (95% CI)**	**Precision (95% CI)**	**F1-score**
Dataset 1 (test)	0.811 (0.719-0.878)	0.809 (0.671-0.903)	0.813 (0.676-0.908)	0.809 (0.671-0.903)	0.809
Dataset 2 (test)	0.704 (0.593-0.798)	0.590 (0.429-0.734)	0.810 (0.663-0.904)	0.742 (0.567-0.865)	0.657
CI: Confidence interval					

Discussion

This study explored the role of a hybrid model combining deep learning and machine learning as a supportive non-invasive decision-support tool that may assist radiologists in the differential diagnosis of IGM and malignant NME on breast MRI. Using DCE-MRI contrast-enhanced and subtraction images, a hybrid deep learning-machine learning model achieved an AUC of 0.870 and an accuracy of 0.811 on the contrast-enhanced images, with lower performance observed on subtraction images (AUC 0.807; accuracy 0.704). Overall, these findings support the potential of DCE-MRI-based hybrid deep-learning and machine-learning approaches for non-invasive differentiation between IGM and malignant NME. The model performed exceptionally well during training, but there was a drop in performance when it was tested on an independent dataset. This difference could be due to overfitting in the model, which is a common problem when using small datasets to train deep learning models.

Breast DCE-MRI plays a central role in detecting lesions presenting as NME, where enhancement morphology and distribution patterns are critical for evaluation (12,13). However, differentiating IGM from malignancy remains challenging, particularly when lesions present as NME, due to considerable imaging overlap (14-16). Soylu Boy (17) reported that although certain MRI features may help in differentiation, the specificity of NME remains limited, and histopathological confirmation is often required.

This diagnostic uncertainty has led to interest in developing artificial intelligence-based classification approaches that aim to improve diagnostic accuracy and reduce dependency on invasive procedures. For example, Zhou et al. (18) demonstrated that a deep learning model achieved diagnostic performance comparable to that of radiologists in differentiating inflammatory breast conditions from malignancy on ultrasound imaging, highlighting the potential of AI as a supportive decision-making tool in complex breast imaging scenarios.

Only a limited number of studies have directly addressed the classification of IGM and malignant NMEs using artificial intelligence applied to DCE-MRI. In one of these studies, Kayadibi et al. (19) investigated the differentiation of IGM and malignant NME using machine learning-based approaches. In this two-center study of 178 patients with NME on breast MRI (69 IGM and 109 breast cancer cases), the authors evaluated clinical models, radiomics models, and combined clinical-radiomics models. Compared with radiologists' interpretation (AUC in training, 0.740; in testing, 0.737), the combined clinical-radiomics model achieved the highest diagnostic performance (AUC 0.979 in training and 0.942 in testing). The study demonstrated that integrating radiomics features with clinical parameters significantly improves the discrimination between IGM and malignant NME, highlighting the complementary role of machine learning in radiological assessment.

Unlike our study, Kayadibi et al. (19) incorporated multiple MRI sequences, including T2W imaging, apparent diffusion coefficient maps, and DCE-MRI. Although we focused on DCE-MRI sequences, to enhance variability and improve generalizability, we also included subtracted images, which are commonly used in routine clinical practice. Another key distinction of our study is the use of deep learning-based image analysis rather than radiomics-based feature extraction, which offers a more practical and automated framework for addressing the same clinical question.

Study Limitations

Our study has some limitations. First, this was a retrospective, single-center study with a limited number of patients; multiple slices were obtained from each patient to expand the dataset. The potential risk of data leakage and overfitting arising from multiple samples from a single patient was mitigated by employing a sequential dataset-splitting strategy. Second, the NME group was heterogeneous with respect to pathological subtypes, and future studies could investigate each subtype in greater

detail. Third, only the second dynamic phase and its subtracted images were analyzed, as these are commonly used in routine practice; studies incorporating a full MRI protocol (e.g., T2W imaging, diffusion-weighted imaging, or the full dynamic series) may provide more comprehensive insights. Fourth, this study did not include an external test set. Fifth, although the numbers of images in the IGM and NME groups were similar, they were not equal due to stringent exclusion criteria. No direct comparison with experienced radiologists was made. Another significant limitation of this study is the lack of an external validation cohort. To validate the generalizability and robustness of the proposed model, future multicenter studies utilizing independent external datasets are necessary.

Despite these limitations, this study has several strengths, including the use of pathologically confirmed cases and the incorporation of both subtracted and non-subtracted DCE-MRI images for model development. In addition, a practical and reproducible deep learning approach was used for image feature extraction. Furthermore, the use of cross-validation and independent internal testing strengthens the methodological framework and supports the reliability of the reported performance metrics.

Conclusion

In the present study, differentiation between IGM and malignant NME on breast MRI was achieved using a hybrid deep learning-based framework. The findings indicate that deep learning–assisted analysis of DCE-MRI may provide additional support for clinical decision-making in the differential diagnosis of IGM and malignant NME. Such approaches have the potential to reduce diagnostic uncertainty and may contribute to avoiding unnecessary invasive procedures in clinically equivocal cases. Nevertheless, these models should be regarded as complementary decision-support tools rather than standalone diagnostic systems, and further validation through larger multicenter prospective studies is required before routine clinical implementation.

Ethics

Ethics Committee Approval: The approval was obtained from the Non-Interventional Clinical Research Ethics Committee of Recep Tayyip Erdogan University Faculty of Medicine (approval no: 2023/107, date: 30.03.2023).

Informed Consent: Written informed consent was obtained from all patients before undergoing breast DCE-MRI.

Footnotes

Authorship Contributions

Concept: F.T., E.K., Design: F.T., E.K., Data Collection or Processing: F.T., E.K., M.N.E., A.A., H.E., N.H.,

Analysis or Interpretation: F.T., E.K., M.N.E., A.A., H.E., N.H., Literature Search: E.K., Writing: F.T., E.K., M.N.E., A.A., H.E., N.H.

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

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

Use of AI for Writing Assistance: During the revision of this work, the author(s) utilized ChatGPT 5.2 (OpenAI) solely for language editing and grammar correction. All outputs were carefully reviewed and revised by the authors to ensure accuracy and consistency. The authors take full responsibility for the content of the publication.

References

1. Han Y, Shi L, Zhang Y. Efficacy and tolerability of methotrexate for idiopathic granulomatous mastitis: a systematic review and meta-analysis. Breast J. 2026;2026:6710172.
2. Vercoe J, Sedaghat N, Brennan ME. Intralesional steroid injections for management of granulomatous mastitis: a systematic review of treatment protocols and clinical outcomes. Breast J. 2025;2025:2592366.
3. Dilaveri C, Degnim A, Lee C, DeSimone D, Moldoveanu D, Ghosh K. Idiopathic granulomatous mastitis. Breast J. 2024;2024:6693720.
4. Yazdipour N, Motamedfar A, Gharibvand MM, Fazelinezhad Z, Farhadi E. Diagnostic accuracy of breast ultrasonography parameters in idiopathic granulomatous mastitis patients: a cross-sectional study. Maedica (Bucur). 2025;20:721-8.
5. Shanbhag NM, Ameri MA, Shanbhag SN, Anandan N, Balaraj K, Bin Sumaida A. Diagnostic challenges and insights into granulomatous mastitis: a systematic review. Cureus. 2024;16:e75733.
6. Karami MY, Amestejani M, Zangouri V, et al. The effectiveness of local steroid injection for the treatment of breast-limited idiopathic granulomatous mastitis: a randomized controlled clinical trial study. Pol Przegl Chir. 2025;97:35-43.
7. Lyu S, Wang B, Xie T, et al. Multiparametric MRI for differentiating idiopathic granulomatous mastitis from invasive breast cancer: improving radiologists' diagnostic accuracy. Eur J Radiol. 2025;184:111958.
8. Yin L, Wei X, Zhang Q, et al. Multimodal ultrasound assessment of mass and non-mass enhancements by MRI: diagnostic accuracy in idiopathic granulomatous mastitis and breast cancer. Breast. 2024;78:103797.

9. Stone M. Cross-validation: a review. Series Statistics. 1978;9:127-39.
10. Sunnetci, K. M., Ulukaya, S. Alkan, A. Periodontal bone loss detection based on hybrid deep learning and machine learning models with a user-friendly application. Biomedical Signal Processing and Control. 2022;77:103844.
11. MathWorks. Choose classifier options – neural network classifiers. MATLAB Documentation. Natick (MA): The MathWorks Inc; 2025. Available from: https://www.mathworks.com/help/stats/choose-classifier-options.html. Accessed: 2025 May 10.
12. Newell D, Nie K, Chen JH, et al. Selection of diagnostic features on breast MRI to differentiate between malignant and benign lesions using computer-aided diagnosis: differences in lesions presenting as mass and non-mass-like enhancement. Eur Radiol. 2010;20:771-81.
13. Zhao Q, Xie T, Fu C, et al. Differentiation between idiopathic granulomatous mastitis and invasive breast carcinoma, both presenting with non-mass enhancement without rim-enhanced masses: the value of whole-lesion histogram and texture analysis using apparent diffusion coefficient. Eur J Radiol. 2020;123:108782.
14. Qu N, Luo Y, Yu T. Differentiation between clinically noninflammatory granulomatous lobular mastitis and noncalcified ductal carcinoma in situ using dynamic contrast-enhanced magnetic resonance imaging. Breast Care. 2020;15:619-27.
15. Wang L, Wang D, Fei X, et al. A rim-enhanced mass with central cystic changes on MR imaging: how to distinguish breast cancer from inflammatory breast diseases? PLoS One. 2014;9:e90355.
16. Azzam MI, Alnaimat F, Al-Nazer MW, et al. Idiopathic granulomatous mastitis: clinical, histopathological, and radiological characteristics and management approaches. Rheumatol Int. 2023;43:1859-69.
17. Soylu Boy FN. MR imaging evaluation of the volume changes and the signs of deformation in the breasts with granulomatous mastitis. Bosphorus Med J. 2022;9:127-131.
18. Zhou Y, Feng BJ, Yue WW, et al. Differentiating non-lactating mastitis and malignant breast tumors by deep-learning based AI automatic classification system: a preliminary study. Front Oncol. 2022;12:997306.
19. Kayadibi Y, Saracoglu MS, Kurt SA, et al. Differentiation of malignancy and idiopathic granulomatous mastitis presenting as non-mass lesions on MRI: radiological, clinical, radiomics, and clinical-radiomics models. Acad Radiol. 2024;31:3511-23.

Original Article

DOI: 10.4274/haseki.galenos.2026.50479
Med Bull Haseki 2026;64(2):85-91

Evaluation of the Accuracy of ChatGPT-generated Information on Human Papillomavirus: A Physician-based Assessment Study

Halil Cagri Aybal[1], Mehmet Yilmaz[2], Mehmet Duvarcı[3], Esra Kiratli Nalbant[4], Isa Dagli[3], Lutfi Tunc[5]

[1]*University of Health Sciences Türkiye, Gulhane Training and Research Hospital, Clinic of Urology, Ankara, Türkiye*

[2]*MEDICLIN Kraichgau-Klinik, Department of Urology, Bad Rappenau, Germany*

[3]*University of Health Sciences Türkiye, Dr. Abdurrahman Yurtaslan Ankara Oncology Training and Research Hospital, Clinic of Urology, Ankara, Türkiye*

[4]*Ankara Bilkent City Hospital, Clinic of Dermatology, Ankara, Türkiye*

[5]*Acıbadem Ankara Hospital, Clinic of Urology, Ankara, Türkiye*

Abstract

Aim: Artificial intelligence (AI) applications are widely used to identify solutions to patients' problems. This study aims to evaluate the scientific validity of information that patients can access about human papillomavirus (HPV) -related topics using Chat-Generative Pre-Trained Transformer (ChatGPT).

Methods: This study was conducted between July 1 and August 1, 2025. A physician developed a structured set of HPV-related questions. The responses generated by ChatGPT were independently evaluated by three clinicians with clinical experience in HPV management. Each response was rated using a five-point Likert scale based on accuracy and clinical relevance. Inter-rater reliability among reviewers was assessed using Cohen's kappa statistic.

Results: The mean scores given by the reviewers for evaluating the accuracy of ChatGPT's answers to HPV-related questions were 4.9±0.3, 4.75±0.44, and 4.75±0.55, respectively. The percentages of correct scores assigned to ChatGPT by the reviewers were 90%, 75%, and 80%, respectively. The approximately equal percentages of correct and incorrect scores were 0, 0, and 5, respectively. The percentages of nearly correct scores were 10, 25, and 15, respectively.

Conclusion: Chat-Generative Pre-Trained Transformer 4.0 demonstrated high efficacy in providing general information regarding HPV, with an 81.6% accuracy rate and a 90% near-accuracy rate. Incorporating AI tools into the facilitation of patient access to information could enhance learning processes. However, it is essential that these tools be continuously refined and utilized to complement rather than substitute for the critical judgment of medical professionals and patients.

Keywords: Artificial Intelligence, humans, human papillomavirus viruses, patient education as a topic

Introduction

Human papillomavirus (HPV) infection is one of the most widespread sexually transmitted diseases in the world, affecting both men and women (1). Human papillomavirus is transmitted through skin-to-skin contact (2). Low-risk HPV can cause genital warts, which can negatively affect patients' social, physical, and sexual lives. High-risk types of HPV can cause cancers of the anus, vagina, vulva, penis, mouth, lungs, and throat (3,4). Infections with HPV can also lead to adverse reproductive outcomes, including reduced sperm quality and decreased sperm concentration and motility (5,6).

Corresponding Author: Halil Cagri Aybal, MD, University of Health Sciences Türkiye, Gulhane Training and Research Hospital, Clinic of Urology, Ankara, Türkiye
E-mail: halilcagri@gmail.com **ORCID:** orcid.org/0000-0001-9123-6139
Received: 12.09.2025 **Accepted:** 08.03.2026 **Publication Date:** 27.03.2026

Cite this article as: Aybal HC, Yilmaz M, Duvarcı M, Kiratli Nalbant E, Dagli I, Tunc L. Evaluation of the accuracy of ChatGPT-generated ınformation on human papillomavirus: a physician-based assessment study. Med Bull Haseki. 2026;64(2):85-91

Large language models (LLMs) are natural language processing (NLP) models that use deep learning algorithms to process and generate text in a manner similar to human language. Chat-Generative Pre-Trained Transformer (ChatGPT) is an NLP model developed by OpenAI that was introduced at the end of 2022 (7). Chat-Generative Pre-Trained Transformer can generate highly accurate and appropriate responses for patient education purposes. Its alignment with professional medical guidelines demonstrates its high potential for patient education (8,9). Additionally, ChatGPT can be used by physicians to facilitate research innovation and comprehensive health management, as well as for diagnostic reasoning in some diseases, particularly in areas where rapid information retrieval and analysis are crucial for patient care (10,11).

In the present study, we hypothesized that clinicians with HPV-related clinical experience could evaluate ChatGPT's information about HPV and thereby determine the accuracy of the information that patients can access through ChatGPT. Patients are using chatbots that use artificial intelligence (AI) more and more to obtain health information, but it's still not clear how reliable the information they provide about HPV-related topics is. Consequently, assessing the precision of AI-generated information is crucial for ascertaining the safe application of such tools in patient education.

Materials and Methods

Compliance with Ethical Standards

The study did not require approval by the institutional review board or ethics committee because no patient data was used. Informed consent was not specifically obtained from respondents. However, it was indicated in their survey responses. We adhered to the Strengthening the Reporting of Observational Studies in Epidemiology guidelines for reporting (12).

Study Design

This observational study was conducted from 1 July to 1 August 2025. A physician generated a set of questions. (M.Y) (Table 1). All questions were chosen subjectively to represent each physician's area of expertise, with the aim of providing a robust and balanced overview of the relevant topics. The physician was requested to provide questions with clearly defined, evidence-based answers based on guidelines from the European Association of Urology, the American Urological Association, and the European Academy of Dermatology and Venereology. The physician developed binary (yes/no), descriptive, or multiple-correct-answer questions, all of which had similar difficulty ratings. A Likert scale can be useful for assessing the accuracy, completeness, and reliability of knowledge (13). The physician who created the questions intended to use a predefined Likert scale to assess their accuracy.

The accuracy scale was a 5-point Likert scale, with the following options: (1: completely incorrect; 2: more incorrect than correct; 3: approximately equally correct and incorrect; 4: nearly correct; and 5: completely correct.)

To assess the consistency of all questions related to HPV, a chatbot version 4.0 was used. To ensure consistency, one investigator (H.C.A.) entered all questions into the chatbot, prompting it with the phrase "Please ensure that information is medically accurate and based on current best practices and guidelines for urology, dermatology, obstetrics, and gynecology" and using unconditional prompts for each new chat. Three specialists (two urologists and one dermatologist) who specialize in the medical or surgical treatment of HPV warts in outpatient clinics were invited to assess the answers (Table 1). To avoid potential bias, respondents were instructed not to use the chatbot to screen the questions themselves. Artificial intelligence-generated answers were reviewed by three clinicians using a five-point Likert scale. The study process is summarized in Figure 1.

Statistical Analysis

Statistical analyses were performed using SPSS version 22.0 (IBM Inc., Armonk, NY, USA). Descriptive statistics were used to summarize the data. Continuous variables

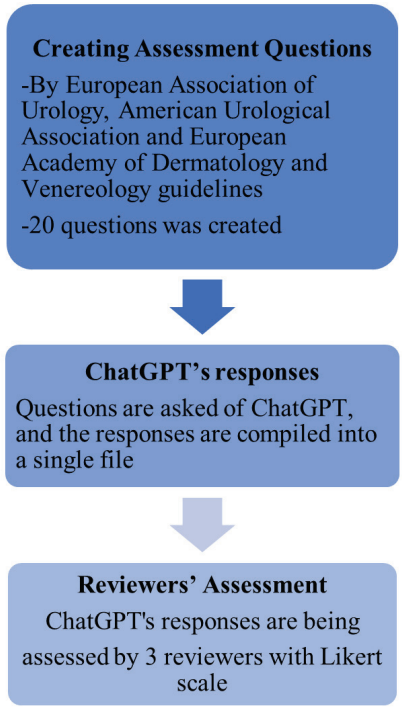

Figure 1. Flowchart of study. Creating assessment questions, ChatGPT's responses and, reviewers' assessment
ChatGPT: Chat-Generative Pre-Trained Transformer

Table 1. Evaluation of HPV related questions

HPV related questions	Answer accuracy point			
	Reviewer 1	Reviewer 2	Reviewer 3	Mean ± SD
1. Which variant of the human papilloma virus is the most commonly detected?	5	5	5	5±0
2. Which variant of human papilloma virus is the most commonly detected at anogenital sites?	5	5	5	5±0
3. Which variant of human papilloma virus is the most commonly detected in anal cytology?	4	4	5	4.33±0
4. Can human papilloma virus be detected in a complete urinalysis?	5	5	5	5±0
5. What is the detection rate of human papilloma virus in urine sample?	5	5	5	5±0
6. Can human papilloma virus be detected in semen?	5	5	5	5±0
7. Does human papilloma virus affect fertility?	5	5	5	5±0
8. What is the frequency of malignant tumour development in human papilloma virus infection?	5	5	5	5±0
9. How often does human papilloma virus infection cause penile cancer?	5	5	5	5±0
10. How often does human papilloma virus infection cause anal cancer?	5	5	5	5±0
11. What is the incidence of oral human papilloma virus?	5	4	3	4±0.81
12. What are the risk factors for human papilloma virus infection?	5	5	4	4.66±0.47
13. How can human papillomavirus be prevented?	5	5	5	5±0
14. How is human papilloma virus transmitted to humans?	5	4	4	4.33±0.47
15. How long does it take for the virus to clear from the body?	4	4	5	4.33±0.47
16. Which variant of human papilloma virus has the shortest clearance time?	5	5	5	5±0
17. Is there an approved test for diagnosing human papilloma virus?	5	5	5	5±0
18. How is human papilloma virus diagnosed?	5	5	5	5±0
19. What is the treatment for human papilloma virus?	5	4	4	4.33±0.47
20. What are the recurrence rates for self-treatment and treatment administered by a physician?	5	5	5	5±0
Mean ± SD	4.9±0.3	4.75±0.44	4.75±0.55	
21. Are these questions suitable for assessing knowledge of HPV?	5	5	5	5±0

1: completely incorrect; 2: more incorrect than correct; 3: approximately equal correct and incorrect; 4: nearly correct; and 5: completely correct
HPV: Human papillomavirus, SD: Standart deviation

were expressed as mean ± standard deviation. Categorical variables were described in terms of frequency and percentage. Cohen's kappa is used to assess inter-rater reliability (IRR). Cohen's kappa correlation values were described as poor agreement: 0.00, slight agreement: 0.00-0.20, fair agreement: 0.21-0.40, moderate agreement: 0.41-0.60, substantial agreement: 0.61-0.80, and almost perfect agreement: 0.81-1.00 (14). A $p<0.05$ was considered significant.

Results

70% of the respondents confirmed that the answers to questions 1, 2, 4-10, 13, 16-18, and 20 were entirely correct. One reviewer confirmed that the answers to questions 3, 14, 15, and 19 were entirely correct, while the other two reviewers confirmed that they were nearly correct. Two reviewers confirmed that the answer to question number 12 was entirely correct. One reviewer confirmed that it was nearly correct. Three reviewers assessed the answer to question 11 as follows: one considered it entirely correct, one considered it nearly correct, and one considered it approximately equally correct and incorrect.

The mean scores given by the reviewers for assessing the accuracy of ChatGPT's answers to HPV-related questions were 4.9±0.3, 4.75±0.44, and 4.75±0.55 respectively. The reviewers' evaluations are presented in Table 1.

The correct score percentages given by the reviewers to ChatGPT were 90, 75, and 80, respectively. The

approximately equal correct and incorrect score percentages were 0, 0, and 5, respectively. The nearly correct score percentages were 10, 25, and 15, respectively. IRR values between reviewer 1 vs. 2, reviewer 1 vs. 3, and reviewer 2 vs. 3 were 0.500 (p=0.01), -0.132 (p=0.473), and 0.448 (p=0.021), respectively. Although no statistically significant agreement was found between the first and third reviewers regarding the accuracy of ChatGPT's responses, statistically significant agreement was observed among the other reviewers. The reviewers' IRR values are shown in Table 2.

Discussion

Large language models can be utilized by humans to identify health issues and direct the treatment process. Chat-Generative Pre-Trained Transformer, developed by OpenAI, is one of the LLMs used for consulting on health issues. This study was designed to determine how accurately ChatGPT could answer questions posed by patients with HPV attending the outpatient clinic and to inform patients about the HPV treatment process.

According to (Surveillance, Epidemiology, and End Results) statistics, the incidence of anal carcinoma increased by an average of 2.2% per year between 2013 and 2022. During this period, it accounted for 0.5% of all new cancer diagnoses in the United States (15). Anal cancer incidence is rising, and predominantly HPV type 16 causes a high-grade squamous intraepithelial lesion (16). High-grade squamous intraepithelial lesion is the precursor lesion of anal squamous cell carcinoma (SCC). It is caused by the uncontrolled growth of squamous epithelial cells in the perianal area or the anal canal transformation zone and is a direct result of an HPV infection. These premalignant lesions may develop into anal SCC if treatment is not received (17). Approximately 88% of anal cancer cases test positive for HPV DNA, indicating a strong association between HPV and anal cancer, second only to cervical cancer. Consequently, HPV has a significant impact on the development of anal cancer (18). Although ChatGPT stated that HPV-16 had a 70% likelihood of causing anal cancer, its report that this strain is the most frequently found in anal cytology was accurate. The use of ChatGPT for medical research regarding the ratio of anal cancer cases associated with HPV 16 is not substantiated by empirical data.

A self-sampling strategy combining HPV detection in urine samples with accessible polymerase chain reaction (PCR) tools was developed as an alternative to cervical swab-based HPV screening to improve participation rates. The PCR kit can detect 14 types of HPV, including HPV-52, HPV-16, and HPV-18, in cervical and urine samples. Urine samples show promise in terms of their accuracy for HPV detection, which could increase cervical cancer screening (19). Our study revealed that ChatGPT correctly identified certain facts relating to cervical cancer screening tests, advanced urine sample options, and the PCR requirement.

The presence of HPV in sperm is associated with male infertility, as indicated by an elevated risk of oligozoospermia and asthenospermia (20). HPV in women has potentially been caused by cervical or tubal factor infertility. However, a scoping review concluded that any studies investigating HPV infection in relation to female fertility had not been conducted (21). Chat-Generative Pre-Trained Transformer has conducted a thorough review of the extant literature on HPV-related infertility in women and men, offering a comprehensive interpretation of the subject. Furthermore, it provides more comprehensive answers to the question by offering information on fertilization, implantation, the probability of pregnancy, and the transmission risk to the partner or the fetus.

Head and neck SCCs (HNSCCs) emerge from the mucosal epithelium of the oral cavity, larynx, and pharynx. The primary risk factor for HNSCCs of the larynx and oral cavity is smoking. Oropharyngeal tumors are increasingly being associated with a history of infection with carcinogenic strains of the HPV, particularly HPV-16. To a lesser extent, this association has also been observed with HPV-18 and other strains (22). The overall prevalence of oropharyngeal HPV in healthy adults in the United States of America and in Europe was reported to be between

Table 2. Inter-rater reliability assessment of reviewers

Points	Reviewer 1 (%)	Reviewer 2 (%)	Reviewer 3 (%)	IRR Reviewer 1 vs. 2*	IRR Reviewer 1 vs. 3*	IRR Reviewer 2 vs. 3*
3	0 (0)	0 (0)	1 (5)	0.500	-0.132	0.448
4	2 (10)	5 (25)	3 (15)			
5	18 (90)	15 (75)	16 (80)			
p-value				**0.01**	0.473	**0.021**

*Cohen's Kappa value
IRR: Inter-rater reliability

3.6% and 6.8% in females and between 6.6% and 15.0% in males (23). The highest oral HPV prevalence was described in South America and the lowest was described in Asia, at 12.4% and 2.6%, respectively (24). Chat-Generative Pre-Trained Transformer has stated that the prevalence of oral HPV is around 7%, oral HPV-16 around 1%, and HPV-related infections around 1% per year. This information does not align with the existing literature on the subject. Furthermore, ChatGPT has stated that HPV-related infections may be temporary and may clear naturally within six to twelve months. This information conflicts with what is published in the literature. (25).

Men who have sex with men, HIV positive status, sexual history (number of lifetime sex partners, number of recent oral or anal sex partners, age of sexual debut), and smoking are associated with adult HPV infection (26). However, ChatGPT noted that lack of male circumcision, other sexually transmitted infections (STIs), and long-term oral contraceptive use are risk factors. The use of oral contraceptives showed an independent association with HPV16-18 infection rates (27). Nevertheless, no clear association exists between HPV infection and circumcision. Additionally, a history of STIs can affect penile, cervical, or vaginal infections.

HPV is transmitted through sexual contact; from mother to fetus; through skin contact (e.g., via the hands or contact with underwear or other inanimate objects); and by high-temperature evaporation treatment. Visible warts can be treated with physical therapy or surgery. The recurrence rate of subclinical infections caused by warts up to 1 cm can be reduced by applying laser therapy, cryotherapy, topical imiquimod, and photodynamic therapy (25). Chat-Generative Pre-Trained Transformer stated that autoinoculation, which is considered a form of skin-to-skin transmission, is a distinct cause of transmission. The information provided by ChatGPT is accurate but insufficient. Furthermore, it noted that indirect contact was very rare and clinically insignificant. Additionally, ChatGPT noted that there is no treatment for asymptomatic HPV infections and that vaccination and surveillance are sufficient. These recommendations do not align with the existing literature on the subject.

Vaccination is the most practical and cost-effective method for avoiding HPV-related health issues. HPV vaccination has been shown to prevent more than 90% of HPV-related cancers (28). In a different study, HPV vaccination was found to be associated with a reduced incidence of several types of cancer among females aged between 9 and 26 years (29). Chat-Generative Pre-Trained Transformer recommended vaccination by age and risk group; this recommendation aligns with the literature on the subject. Furthermore, it emphasized the significance of HPV vaccination and the adoption of safe sex practices, including limiting the number of sexual partners, using condoms, and enhancing immune function through smoking cessation.

The clearance time for HPV has been referred to as being less than 4 weeks (±4 weeks) in the short term and 12 months (±6 months) in the long term (25). Albero et al. (30) concluded that the period during which the HPV virus was cleared ranged from 1.3 to 42.1 months. According to ChatGPT, the median clearance time for HPV is between 6 and 18 months. HPV infections usually clear spontaneously within one to two years in immunocompetent individuals. It stated that persistence may last 12-24 months or longer, particularly for high-risk types.

Study Limitations

This study was conducted with only the ChatGPT 4.0 model, and other AI-based models were not included in the evaluation. The accuracy of the responses was assessed by a limited number of physicians, thereby restricting the diversity of perspectives. Only three reviewers, physicians from overlapping specialties, were included. The inclusion of a larger group of experts could have increased the reliability of the evaluations. Future studies involving multiple AI models and a broader range of physicians may strengthen the generalizability of the findings, particularly by including diverse specialties and practice settings to ensure a more comprehensive evaluation of the AI's effectiveness across different patient populations. Despite these limitations, the study will enable patients to obtain preliminary information based on highly accurate data evaluated by reviewers specified by ChatGPT during the interval between seeking medical attention and visiting a physician.

Conclusion

Chat-Generative Pre-Trained Transformer 4.0 has demonstrated high efficacy in providing general information about HPV, with an 81.6% accuracy rate and a 90% near-accuracy rate. Despite the evident potential demonstrated by ChatGPT, these findings should be regarded as preliminary indications of its promise, particularly in research related to human health. Furthermore, these results should not be construed as validation of ChatGPT's clinical adequacy. Incorporating AI tools into the facilitation of patient access to information could enhance learning processes. However, it is essential that these tools be continuously refined and utilized to complement, rather than substitute, the critical judgment of medical professionals and patients.

Ethics

Ethics Committee Approval: The study did not require approval from the institutional review board or ethics committee because no patient data was used.

Informed Consent: Since this study is based on responses generated by artificial intelligence, patient consent was not required.

Footnotes

Authorship Contributions

Concept: H.C.A., Design: L.T., Data Collection or Processing: M.Y., Analysis or Interpretation: H.C.A., Literature Search: M.D., E.K.N., I.D., Writing: H.C.A., M.Y.

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

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

References:

1. Khan I, Harshithkumar R, More A, Mukherjee A. Human papilloma virus: an unraveled enigma of universal burden of malignancies. Pathogens. 2023;12:564.
2. Szymonowicz KA, Chen J. Biological and clinical aspects of HPV-related cancers. Cancer Biol Med. 2020;17:864-78.
3. Dominiak-Felden G, Cohet C, Atrux-Tallau S, Gilet H, Tristram A, Fiander A. Impact of human papillomavirus-related genital diseases on quality of life and psychosocial wellbeing: results of an observational, health-related quality of life study in the UK. BMC Public Health. 2013;13:1065.
4. Kohli M, Bunker CB, Kravvas G. Human papillomavirus: an update. Clin Dermatol. 2026;44:54-66.
5. Cao X, Wei R, Zhang X, Zhou J, Lou J, Cui Y. Impact of human papillomavirus infection in semen on sperm progressive motility in infertile men: a systematic review and meta-analysis. Reprod Biol Endocrinol. 2020;18:38.
6. Wang QH, Ye JJ, Chen ZY, et al. Current risk factors for male infertility and semen parameters: an umbrella review of systematic reviews and meta-analyses. Asian J Androl. 2026.
7. Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of large language models for medical education and knowledge assessment. JMIR Med Educ. 2023;9:e45312. Erratum in: JMIR Med Educ. 2024;10:e57594.
8. Emile SH, Horesh N, Garoufalia Z, Gefen R, Boutros M, Wexner SD. Assessment of the utility of artificial intelligence-based chatbots in patient education: a systematic review and meta-analysis. The American Surgeon™. 2026;92:258-69.
9. Pandya S, Alessandri Bonetti M, Liu HY, Jeong T, Ziembicki JA, Egro FM. Burn patient education in the modern age: a comparative analysis of chatgpt and google performance answering common questions on burn injury and management. J Burn Care Res. 2025;46:533-41.
10. Lanzafame LRM, Gulli C, Mazziotti S, et al. Chatbots in radiology: current applications, limitations and future directions of chatgpt in medical imaging. Diagnostics (Basel). 2025;15:1635.
11. Zhu Y, Luo D, Shen X, et al. Application of ChatGPT-based artificial intelligence in the diagnosis and management of polycystic ovary syndrome. BMC Med Inform Decis Mak. 2025;25:271.
12. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. J Clin Epidemiol. 2008;61:344-9.
13. Molena KF, Macedo AP, Ijaz A, et al. Assessing the accuracy, completeness, and reliability of artificial intelligence-generated responses in dentistry: a pilot study evaluating the ChatGPT model. Cureus. 2024;16:e65658.
14. Fageeh HI, Fageeh HN, Bhati AK, et al. Assessing the reliability of miller's classification and cairo's classification in classifying gingival recession defects: a comparison study. Medicina (Kaunas). 2024;60:205.
15. Surveillance, Epidemiology, and End Results Program. Cancer Stat Facts: Anal cancer. 2025. Available from: https://seer.cancer.gov/statfacts/html/anus.html. Accessed March 3, 2025.
16. Rozemeijer K, Dias Gonçalves Lima F, Kuyvenhoven EJ, et al. Swab-based anal cancer screening in men living with HIV: projected outcomes for different screening algorithms. Int J Cancer. 2025;157:2259-68.
17. Araujo ROC, Valadão M, Silva JADDCE, et al. Implementation of a screening program for high-grade anal dysplasia in high-risk patients in a tertiary cancer center. J Surg Oncol. 2025;131:151-9.
18. Ebrahimi F, Rasizadeh R, Jafari S, Baghi HB. Prevalence of HPV in anal cancer: exploring the role of infection and inflammation. Infect Agent Cancer. 2024;19:63.
19. Intan NS, Utama R, Wulandari D, et al. Urinary detection of high-risk HPV DNA to enhance cervical cancer screening in developing countries. Microbiol Spectr. 2025;13:e0193824.
20. Priam A, Bozec AL, Meireles VD, et al. Human papillomavirus carriage in the semen of men consulting for infertility: prevalence and correlations with sperm characteristics. Asian J Androl. 2025;27:196-203.
21. Kristensen TS, Foldager A, Laursen ASD, Mikkelsen EM. Sexually transmitted infections (Chlamydia trachomatis, genital HSV, and HPV) and female fertility: a scoping review. Sex Reprod Healthc. 2025;43:101067.
22. Johnson DE, Burtness B, Leemans CR, Lui VWY, Bauman JE, Grandis JR. Head and neck squamous cell carcinoma. Nat Rev Dis Primers. 2020;6:92. Erratum in: Nat Rev Dis Primers. 2023;9:4.
23. Alemany L, Felsher M, Giuliano AR, et al. Oral human papillomavirus (HPV) prevalence and genotyping among healthy adult populations in the United States and Europe: results from the PROGRESS (PRevalence of Oral hpv infection, a Global aSSessment) study. EClinicalMedicine. 2025;79:103018.

24. Yu S, Zhu Y, He H, et al. Prevalence and risk factors of oral human papillomavirus infection among 4212 healthy adults in Hebei, China. BMC Infect Dis. 2023;23:773.

25. Zhu P, Qi RQ, Yang Y, et al. Clinical guideline for the diagnosis and treatment of cutaneous warts (2022). J Evid Based Med. 2022;15:284-301.

26. Del Pino M, Vorsters A, Joura EA, et al. Risk factors for human papillomavirus infection and disease: a targeted literature summary. J Med Virol. 2024;96:e29420.

27. Li L, Wu H, Chen Y, et al. Influence of pre-vaccination HPV status on vaccine effectiveness among chinese women: a multicenter cross-sectional study. Cancer Rep (Hoboken). 2025;8:e70294.

28. Stuart R, Theopold N, Miall N, et al. The role of HPV single-dose vaccination in expanding access in GAVI-supported countries during a period of supply constraints. Vaccine. 2026;75:128187.

29. Hung YM, Lin TT, Wang SI, Wu PJ, Chang R, Wei JC. HPV vaccination is associated with lower risk of cancers among females. Am J Med. 2026;139:311-20.e12.

30. Albero G, Castellsagué X, Giuliano AR, Bosch FX. Male circumcision and genital human papillomavirus: a systematic review and meta-analysis. Sex Transm Dis. 2012;39:104-13.

Original Article

DOI: 10.4274/haseki.galenos.2026.97269
Med Bull Haseki 2026;64(2):92-100

Perioperative Effects of Epidural Anesthesia on Hemodynamics, Pain, and Pro-brain Natriuretic Peptide Levels in Vascular Surgery: A Prospective Randomized Study

Mehmet Ilker Sivacigil[1], Leyla Sivacigil[2], Lale Yuceyar[3], Hulya Erolcay[3]

[1]*University of Health Sciences Türkiye, Istanbul Beyoglu Eye Training and Research Hospital, Clinic of Anesthesiology and Reanimation, Istanbul, Türkiye*

[2]*University of Health Sciences Türkiye, Taksim Training and Research Hospital, Clinic of Anesthesiology and Reanimation, Istanbul, Türkiye*

[3]*Istanbul University-Cerrahpasa Faculty of Medicine, Department of Anesthesiology and Reanimation, Istanbul, Türkiye*

Abstract

Aim: Vascular surgery is among the surgical procedures associated with the highest cardiac morbidity and mortality because the surgical stress response is pronounced in individuals with cardiogenic risk factors. The aim of this study is to evaluate the effect of epidural anesthesia on the surgical stress response during major vascular surgery.

Methods: Two groups (epidural group (EG), n=15; control group, n=15) were formed in the cardiovascular surgery operating room by prospective randomization. Hemodynamic fluctuations and visual analog scale (VAS) pain scores were monitored. The primary endpoint was defined as the change in N-terminal-proBNP (NT-proBNP) levels on postoperative day two compared with preoperative measurements. Troponin-I, changes in blood flow, pain scores, and sevoflurane use were secondary endpoints.

Results: Sevoflurane consumption was lower in EG, whereas fluid volume was higher. ≥11 tachycardia attacks were more frequent in CG, and ≥11 decreases in mean arterial pressure were seen (33.3% vs. 0%; p=0.042). VAS was lower in favor of EG (Day-1: 4.33±0.98 vs. 5.27±0.70, p=0.011; Day-2: 1.67±1.13 vs. 2.13±0.64, p=0.041). No difference in NT-proBNP/troponin-I levels was observed between groups.

Conclusion: In vascular surgery, epidural anesthesia can reduce hemodynamic fluctuations and early postoperative pain while limiting the rise in NT-proBNP. Biomarker findings are hypothesis-generating owing to methodological constraints and necessitate validation through studies with larger sample sizes.

Keywords: Anesthesia, epidural, vascular surgical procedures, natriuretic peptide, brain, Troponin I, postoperative pain

Introduction

Vascular surgery is among the surgical procedures with the highest cardiac morbidity and mortality during the perioperative period (1-2). The main reason for poor prognosis in patients with cardiogenic risk factors is hemodynamic changes that develop due to surgical stress. Effective management of the surgical stress response is critical for maintaining hemodynamic stability (1,3). Specific biomarkers are needed to predict cardiac mortality and morbidity, especially in high-risk patient groups (4,5).

Natriuretic peptides are hormones involved in the regulation of blood pressure and blood volume. Pro-B-type natriuretic peptide (proBNP) is released into the circulation in response to increased ventricular wall stress and pressure load; therefore, it is sensitive to

Corresponding Author: Leyla Sivacigil, MD, University of Health Sciences Türkiye, Taksim Training and Research Hospital, Clinic of Anesthesiology and Reanimation, Istanbul, Türkiye
E-mail: veralleyla@hotmail.com **ORCID:** orcid.org/0000-0001-6375-5417
Received: 28.10.2025 **Accepted:** 26.02.2026 **Publication Date:** 27.03.2026

Cite this article as: Sivacigil MI, Sivacigil L, Yuceyar L, Erolcay H. Perioperative effects of epidural anesthesia on hemodynamics, pain, and pro-brain natriuretic peptide levels in vascular surgery: a prospective randomized study. Med Bull Haseki. 2026;64(2):92-100

detecting perioperative cardiac stress and the risk of adverse events (4,6). Cardiac troponin I (cTnI) is a specific marker of myocardial damage, and its increase is linearly associated with poor prognosis (5). Since N-terminal-proBNP (NT-proBNP) reflects hemodynamic load and cTnI reflects necrosis, evaluating these two markers together may allow for more comprehensive monitoring of cardiac risk associated with surgical stress (4,7). We hypothesized that epidural anesthesia, when added to general anesthesia for major vascular surgery, provides effective pain control in the perioperative and postoperative periods, partially suppresses the surgical stress response, and reduces cardiac risk by lowering myocardial oxygen demand through sympathetic blockade, peripheral vasodilation, and increased coronary and collateral blood flow (7,8). Based on this rationale, the primary endpoint was defined as the change in NT-proBNP levels from the day before surgery to the day after surgery. Troponin I, the number of blood-flow changes, and visual analog scale (VAS) pain scores were secondary endpoints.

Specifically, prospective randomized data assessing the potential influence of the dynamics of NT-proBNP and troponin-I in relation to hemodynamic variations and pain scores are limited. Few prospective studies have examined the effect of epidural anesthesia on the dynamics of cardiac biomarkers in patients undergoing vascular surgery (4,7,9). The aim of this study is to evaluate the effect of epidural anesthesia on the surgical stress response in major vascular surgery by measuring cardiac biomarkers, hemodynamic variables, and pain scores. This will enable the monitoring of the frequency of surgical stress responses and related cardiac complications in the perioperative and postoperative periods and will contribute to maintaining hemodynamic stability.

Materials and Methods

Compliance With Ethical Standards

The study was conducted in the Cardiovascular Surgery operating room and was approved by the Istanbul University-Cerrahpasa Faculty of Medicine Ethics Committee (approval no.: 43423, date: 01.12.2011), in accordance with the Declaration of Helsinki (10). All patients were provided with written and verbal information, and written consent was obtained. The supply of NT-proBNP and troponin I kits was provided by the Scientific Research Projects (BAP) Unit of Istanbul University-Cerrahpasa Faculty of Medicine University. Kits from Biomedica Slovakia and Diagnostic Automation Inc. were used for NT-proBNP and troponin I, respectively.

Study Design and Population

30 patients undergoing vascular surgery [American Society of Anesthesiologists (ASA) II-III] were randomly allocated to the study groups using a sealed opaque envelope randomization technique: the epidural group (EG), which received epidural anesthesia, and the control group (CG). Patient demographics (height, body weight, age, gender, and ASA status) were recorded. The randomization process and analysis sets are presented in Figure 1.

Exclusion Criteria

Liver dysfunction (alanine aminotransferase/aspartate aminotransferase >40), kidney failure, New York Heart Association (NYHA) IV heart failure, unstable angina, second-third degree atrioventricular (AV) block, pregnancy, glucocorticoid use, hyperthyroidism/hypothyroidism or thyroid hormone use, epidural contraindication (bleeding diathesis, hypovolemic shock, local infection, vertebral deformity, arthritis, osteoporosis, back pain, hypotension, increased intracranial pressure, previous laminectomy, drug allergy, back trauma, cauda equina syndrome).

Anesthesia Management and Monitoring

Following preoperative assessment, electrocardiography (ECG), peripheral oxygen saturation (SpO_2), and arterial blood pressure were monitored; an intravenous cannula and a urinary catheter were inserted; premedication with 0.03 $mg \cdot kg^{-1}$ midazolam was administered, and an appropriate crystalloid infusion was initiated. Invasive arterial pressure was continuously monitored with radial artery catheterization. During service follow-ups throughout the surgery, 20% increases or decreases in heart rate (HR) and mean arterial pressure (MAP) relative to reference values were recorded. HR <50 bpm (bradycardia), a 20% increase in HR (tachycardia), a 20% decrease in MAP (hypotension), and a 20% increase in MAP (hypertension) were considered. Standard algorithms were applied to manage hypotension, bradycardia, tachycardia, and hypertension (sevoflurane adjustment; 5 $mL \cdot kg^{-1}$ crystalloid bolus; 5 mg ephedrine; 0.5 mg atropine; 1 $\mu g \cdot kg^{-1}$ fentanyl).

For epidural anesthesia, 5-7 $mg \cdot kg^{-1}$ of balanced electrolyte solution was administered before catheter placement. The epidural catheter was inserted using the "loss of resistance" method at the L1-L2 or L2-L3 interspace; the block level was assessed using a pinprick after a 3-mL test dose of lidocaine. A preoperative 10-mL bolus of 0.5% bupivacaine +1 mg of morphine was administered epidurally.

Pain Management

For the first 48 hours postoperatively, epidural patient-controlled analgesia (PCA) (100 mL of 0.05 bupivacaine +10 mg morphine; 6 $mL \cdot s^{-1}$ infusion, 4 mL bolus, 30 min lock) and IV-PCA (morphine 1 $mg \cdot mL^{-1}$; 2 mg bolus; 15 min lock) were administered to the CG.

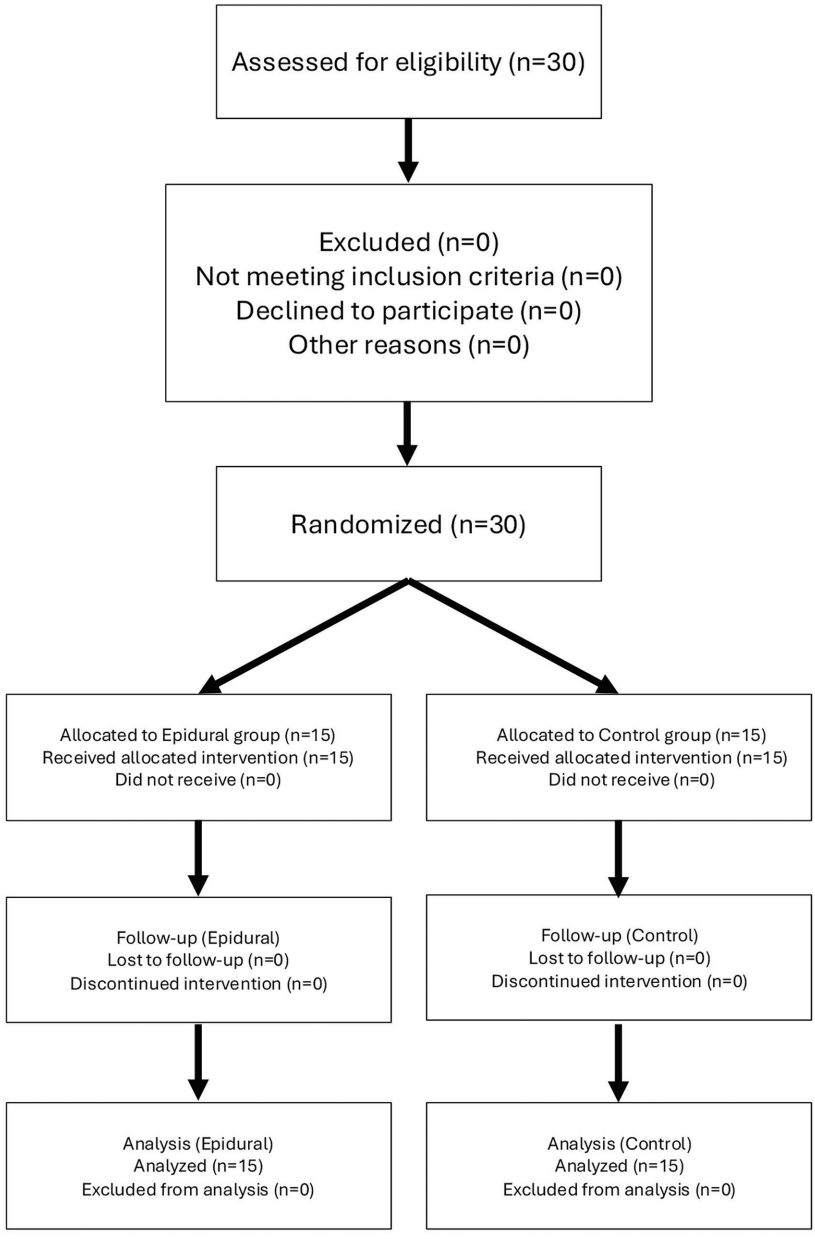

Figure 1. CONSORT flow diagram of patient enrollment, randomization, follow-up, and analysis

The main goal of this study was to determine how NT-proBNP levels changed on the second postoperative day compared with the day before surgery. High-sensitivity cardiac troponin (hs-cTn) was used to assess myocardial injury in the study. This study was conducted to evaluate myocardial injury in patients undergoing non-cardiac surgery. High-sensitivity cardiac troponin was used to assess myocardial injury. Myocardial injury after non-cardiac surgery (MINS) was defined as a postoperative hs-cTn value exceeding the 99th percentile upper reference limit in the absence of a non-ischemic cause, in accordance with the Fourth Universal Definition of Myocardial Infarction. Patients with NYHA class IV heart failure, unstable angina, and second- or third-degree AV block were excluded from the study. Troponin-I levels, number of hemodynamic fluctuations (HR/MAP criteria), VAS scores, and sevoflurane intake were the secondary endpoints. Follow-up and measurements: cardiac events were assessed on postoperative days 1 and 2, and pain was assessed using the VAS. Major adverse cardiac events (MACE), perioperative mortality, and perioperative MI were addressed through early clinical questioning and routine

assessments. In this study, clinical cardiac endpoints (e.g., MACE and mortality) were not systematically monitored as primary or secondary endpoints; instead, they were only addressed through early clinical questioning and routine assessments. Blood samples and ECGs were taken preoperatively, after extubation, and on the 1st and 2nd postoperative days. Plasma samples were centrifuged at 4000 rpm for 10 minutes and stored at -80 °C; NT-proBNP and troponin I measurements were performed after the patient enrollment was completed.

Statistical Analysis

Sample size was determined by one-way ANOVA based on pilot data (α=0.05, power =80%, effect size =0.5), which required at least 17 patients per group. Continuous variables were tested for normality and expressed as mean ± standard deviation or median (minimum-maximum), as appropriate. Categorical variables were presented as numbers and percentages. Between-group comparisons were performed using the independent samples t-test or Mann-Whitney U test, and categorical variables were analyzed using the chi-square or Fisher's exact test. Effect sizes were calculated using Cohen's d (0.2 small, 0.5 medium, and 0.8 large). Changes over time in MAP, HR and VAS were evaluated using the Friedman test with Bonferroni-corrected post-hoc comparisons. Correlations were assessed using Spearman's correlation coefficient. Analyses were conducted using NCSS 2007 and PASS 2008 software, and p<0.05 was considered statistically significant.

Results

The study was conducted in a total of 30 cases undergoing vascular surgery in the operating room of Istanbul University Faculty of Medicine, Department of Cardiovascular Surgery (EG n=15; CG n=15) (Figure 1). The ages of the subjects ranged from 42 to 77 years (59.27±8.78). There was a significant age difference between the groups; the CG was older (EG 55.07±7.43; CG 63.47±8.17; p=0.006). There was no difference in BMI (EG 27.00±2.10; KG 27.67±1.91; p=0.372) and gender distribution (female EG 33.3%; CG 13.3%; p=0.390) (Table 1).

Anesthetic Consumption and Fluids

In the EG, sevoflurane consumption was significantly lower (56.40±13.99 mL vs. 67.40±14.32 mL; p=0.042). The intraoperative fluid volume administered was higher in the EG group (2586.67±1130.02 mL vs. 1870.00±296.29 mL; p=0.030) (Table 2).

Hemodynamic Variables

The rate of "≥11 times" for increased CAD was higher in the CG (EG 13.3%; CG 53.3%; p=0.020). The intergroup difference in MAP increase was not significant. No differences were found among the HR decrease categories. The rate of "≥11 times" occurrence in MAP was significantly higher in the CG group (EG 0%; CG 33.3%; p=0.042) (Table 3).

Table 1. Baseline demographic and anthropometric characteristics of the groups (epidural and control)

	Epidural (n=15)	Control (n=15)	p
	Mean ± SD	Mean ± SD	
Age (years)	55.07±7.43	63.47±8.17	0.006*
BMI (kg/m²)	27.00±2.10	27.67±1.91	0.372
	n (%)	n (%)	
Gender†			0.390
Female	5 (33.3%)	2 (13.3%)	
Male	10 (66.7%)	13 (86.7%)	

*p<0.01. Data are presented as mean ± SD or n (%). Comparisons of age and BMI were made using the independent samples t-test (Mann-Whitney U if the assumption of normality was not met), and for gender†, the chi-square test (Fisher's exact test if necessary). Significance level was p<0.05
SD: Standard deviation, BMI: Body mass index (kg/m²)

Table 2. Sevoflurane consumption and total fluid volume during surgery (epidural and control)

	Epidural (n=15) Mean ± SD	Control (n=15) Mean ± SD	p	Mean difference (E-C)	Pooled SD	Cohen's d
Sevoflurane consumption (mL)	56.40±13.99	67.40±14.32	0.042*	-11.00	14.16	-0.78
Fluid volume (mL)	2586.67±1130.02	1870.00±296.29	0.030*	716.67	826.05	0.87

*p<0.05. Data are presented as mean ± SD. Groups were compared using an independent samples t-test (Mann-Whitney U test when normality was not achieved). Significance level: p<0.05. Mean difference calculated as epidural minus control (E-C). Effect size interpreted according to Cohen's criteria (0.2= small, 0.5= moderate, 0.8= large)
SD: Standard deviation, mL:milliliters

Pain

Visual analog scale was lower in favor of the EG group on day 1 (EG 4.33±0.98; CG 5.27±0.70; p=0.011) and day 2 (EG 1.67±1.13; CG 2.13±0.64; p=0.041); the decrease from day 1 to day 2 was significant in both groups (p=0.001) (Table 4).

Biomarkers

No significant differences were found between groups for preoperative, post-extubation, and postoperative day 1 and day 2 measurements of proBNP and troponin I. This result shows that the primary endpoint does not differ significantly between the groups. In intra-time comparisons, no significant change in proBNP was observed in the EG, whereas a significant increase was detected in the CG after pre-operation (p=0.031), on the 1st postoperative day (p=0.016), and on the 2nd postoperative day (p=0.015). There was no significant effect of time on troponin I in either group (Tables 5 and 6). No significant difference was observed between preoperative NT-proBNP levels and postoperative day 2 levels (p>0.05). High-sensitivity cardiac troponin levels did not exceed the 99th percentile upper reference limit in any patient; therefore, no cases of MINS were identified. No MACE, perioperative myocardial infarction, or deaths were observed.

Table 3. Frequency of intraoperative hemodynamic fluctuations: heart rate (HR) and mean arterial pressure (MAP) changes (epidural and control)

Change	Frequency	Measure	Epidural (n=15)	Control (n=15)	p
Increase	None	HR	3 (20.0%)	1 (6.7%)	0.598
Increase	None	MAP	0 (0.0%)	0 (0.0%)	1.000
Increase	1-10 times	HR	10 (66.7%)	6 (40.0%)	0.143
Increase	1-10 times	MAP	14 (93.3%)	9 (60.0%)	0.080
Increase	≥11 times	HR	2 (13.3%)	8 (53.3%)	0.020*
Increase	≥11 times	MAP	1 (6.7%)	6 (40.0%)	0.080
Decrease	None	HR	2 (13.3%)	0 (0.0%)	0.483
Decrease	None	MAP	2 (13.3%)	0 (0.0%)	0.483
Decrease	1-10 times	HR	11 (73.4%)	10 (66.7%)	1.000
Decrease	1-10 times	MAP	13 (86.7%)	10 (66.7%)	0.390
Decrease	≥11 times	HR	2 (13.3%)	5 (33.3%)	0.390
Decrease	≥11 times	MAP	0 (0.0%)	5 (33.3%)	0.042*

*p<0.05. Data are presented as n (%). "Increase/decrease" indicates a ≥20% increase/decrease in the relevant parameter compared to baseline; "Frequency" rows indicate the number of events (None, 1-10 times, ≥11 times). Groups were compared using the chi-square test (Fisher's exact test if necessary); significance level is p<0.05

Table 4. Postoperative pain levels (VAS) - day 1 and day 2 comparisons (epidural vs. control)

VAS	Epidural (n=15) Mean ± SD (Median)	Control (n=15) Mean ± SD (Median)	p	Cohen's d (Epidural-Control)	Magnitude (|d|)
Day 1	4.33±0.98 (4.00)	5.27±0.70 (5.00)	0.011**	-1.1	Large
Day 2	1.67±1.13 (1.00)	2.13±0.64 (2.00)	0.041*	-0.5	Medium
Day 1/day 2;†	0.001***	0.001***	-	-	-

*p<0.05; **p<0.01; ***p<0.001. Data are presented as mean ± SD (median). For between-group (epidural vs. control) comparisons, an independent samples t-test (Mann-Whitney U test if normality was not achieved) was used. For within-group (day-1 vs. day-2) comparisons, a paired t-test (Wilcoxon signed-rank test if normality was not achieved) was used. Significance level p<0.05. Mean difference calculated as Epidural minus Control (E-C). Effect size interpreted according to Cohen's criteria (0.2= small, 0.5= moderate, 0.8= large)
VAS: Visual analog scale (0-10), SD: Standard deviation

Table 5. Comparison of cardiac biomarkers by time points: proBNP and Troponin-I (epidural and control)						
Time point	**Marker**	**Epidural (Mean ± SD (median)**	**Control (Mean ± SD (median)**	**p**	**Cohen's d (E-C)**	**Magnitude (|d|)**
Pre-op	proBNP (pg/mL)	87.05±115.18 (48.79)	33.53±28.52 (40.38)	0.148	0.638	Medium
	Troponin-I (ng/mL)	0.91±1.42 (0.32)	1.00±1.66 (0.08)	0.967	-0.058	Trivial
Post-extubating	proBNP (pg/mL)	71.53±108.39 (37.02)	54.20±31.80 (57.93)	0.539	0.217	Small
	Troponin-I (ng/mL)	3.24±11.29 (0.01)	1.25±1.67 (0.62)	0.081	0.247	Small
Post-op day 1	proBNP (pg/mL)	42.26±29.05 (32.40)	60.61±47.11 (45.51)	0.389	-0.469	Small
	Troponin-I (ng/mL)	0.69±1.57 (0.08)	1.34±2.28 (0.56)	0.187	-0.332	Small
Post-op day 2	proBNP (pg/mL)	49.27±49.77 (32.95)	99.49±151.35 (52.66)	0.412	-0.446	Small
	Troponin-I (ng/mL)	2.03±5.13 (0.03)	1.01±1.44 (0.76)	0.461	0.271	Small

Data are presented as mean ± SD (median). An independent samples t-test (Mann-Whitney U when normality was not achieved) was used for comparisons between groups at each time point; p<0.05 was considered statistically significant. Mean difference calculated as Epidural minus Control (E-C).Effect size interpreted according to Cohen's criteria (0.2= small, 0.5= moderate, 0.8= large)
proBNP: N-terminal pro-B-type natriuretic peptide (pg/mL), Troponin-I (ng/mL), Pre-op: Before surgery, post-extubating = after extubating, Post-op day 1/2 = post-operative day 1/2; SD: Standard deviation

Table 6. Within-group comparisons over time: paired p-values for proBNP and Troponin-I (epidural and control)				
Comparison time points	**ProBNP epidural (p)**	**ProBNP control (p)**	**Troponin epidural (p)**	**Troponin control (p)**
Pre-op vs. post-extubating	0.570	0.031*	0.575	0.180
Pre-op vs. post-op day 1	0.191	0.016*	0.790	0.594
Pre-op vs. post-op day 2	0.124	0.015*	0.477	0.683
Post-extubating vs. post-op day 1	0.427	0.570	0.594	0.657
Post-extubating vs. post-op day 2	0.820	0.460	0.859	0.865
Post-op day 1 vs. post-op day 2	0.460	0.307	0.722	0.917

*p<0.05. The p-values in the table represent paired comparisons between the two specified time points within each group (epidural, control) (Pre-op ↔ Post-extubating; Pre-op ↔ Post-op Day-1/Day-2; etc.). Analysis was performed using a paired t-test (Wilcoxon signed-rank test when normality was not met); p<0.05 was considered significant.
proBNP: N-terminal pro-B-type natriuretic peptide, Pre-op: Pre-operative, post-extubating = post-extubating, Post-op day 1/2 = post-operative day 1/2

Discussion

In this study, proBNP and troponin-I levels were simultaneously evaluated in vascular surgery patients in the EG who underwent epidural anesthesia to investigate the secondary cardiac effects of epidural anesthesia on the surgical stress response. Current guidelines classify vascular surgery as high cardiac risk and recommend rigorous perioperative and early postoperative surveillance. Reducing the surgical stress response is essential for lowering morbidity and mortality by mitigating adverse effects on the hemodynamic, respiratory, and gastrointestinal systems (1,3).

Pro-B-type natriuretic peptide and troponin-I levels increase during cardiac stress or injury; even slight elevations in troponin have been associated with 30-day mortality (11,12). In vascular patients, elevated preoperative and early postoperative NT-proBNP levels predict MACE (12). This study indicated that the intergroup differences in NT-proBNP and troponin-I levels were not statistically significant; nevertheless, the observed "lesser increase" in EG should be interpreted not as a conclusive effect but as a trend consistent with the prevailing literature. Consequently, the identification of a "vlesser increase" in the EG should be regarded not as conclusive evidence of an effect but as an observation that suggests trends and generates hypotheses. Data demonstrate that high thoracic epidural anesthesia (TEA) can reduce troponin release through its sympatholytic effect; furthermore, it is associated with attenuated changes in NT-proBNP following cardiac surgery (4-7).

At specific intervals, the CG exhibited significantly larger variations in hemodynamic parameters. Thoracic epidural anesthesia/epidural can improve hemodynamic stability by causing vasodilation and lowering HR and blood pressure through sympathetic blockade (13). This study found that tachycardia and hypertensive episodes were less frequent in the EG. Significantly lower VAS scores on postoperative days one and two in EG indicate effective analgesia, which reduces pain-induced sympathetic activation and cardiac strain. The literature supports the idea that epidural

analgesia may reduce VAS scores and opioid use and facilitate early ambulation, which could benefit cardiac function (14,15).

Possible mechanisms include suppression of the sympathetic stress response, reduced increases in catecholamines and cortisol, control of HR and afterload, decreased myocardial O_2 consumption, and coronary vasodilation. Furthermore, reduced opioid use and early mobilization resulting from effective analgesia have been considered factors that could explain the cardioprotective effects of the epidural approach (9). Recent evidence suggests that the incidence of postoperative myocardial infarction after noncardiac surgery may be reduced with epidural anesthesia, and specific studies have shown reductions in mortality rates (9,16).

When interpreting biomarkers, it is important to remember that troponin levels can increase slightly in low-risk individuals and can peak late. It is also important to remember that pre-analytical factors (such as storage and kit conditions) and false positives can affect the results (7). For NT-proBNP, pre-analytical factors such as age and gender effects and the timing and location of sampling are significant (17). A correlation has been observed between intraoperative fluid volume and lung USG B-lines in cesarean patients; however, BNP may not consistently exhibit a direct correlation with the administered fluid (8). Our study did not reveal a significant difference in NT-proBNP and troponin levels between the two groups; however, an increase in NT-proBNP levels was noted in the CG group in the within-group analysis. This observation should be interpreted in the context of the hypothesis that elevated sevoflurane consumption and greater hemodynamic fluctuations in the CG group may have contributed to increased proBNP levels. Research has also reported postoperative BNP elevation in studies comparing spinal and general anesthesia (18).

Pölzl et al. (19) found that high preoperative NT-proBNP levels were associated with 30-day and 5-year mortality in 6,938 patients undergoing cardiac surgery. A high NT-proBNP level at diagnosis and a preoperative NT-proBNP drop below 3000 ng/L were associated with more favorable perioperative outcomes and shorter lengths of stay in the intensive care unit (19). In our study, we found no significant difference in NT-proBNP levels between the groups. We did not include patients with advanced heart failure, advanced kidney failure, or high-grade cardiac arrhythmias in our research. This could explain the lack of a significant difference in NT-proBNP levels. NT-proBNP levels may demonstrate more pronounced differences in patients with advanced heart failure. In this respect, we believe that our study may provide insights for future research.

Laferrière-Langlois et al. (20) examined 33,089 patients who underwent cardiac surgery with epidural anesthesia between 1966 and 2022. They compared the potential benefits of TEA during cardiac surgery in terms of mortality, atrial fibrillation, and pulmonary complications with the risk of epidural hematoma associated with intraoperative heparinization. Across all published studies, no cases of epidural hematoma were reported. The use of epidural anesthesia should be considered a safe treatment option in cardiac surgery (20).

Publications suggesting that epidural anesthesia reduces the perioperative stress response primarily concern thoracic epidural analgesia and cardiac surgery. From this perspective, it is not surprising that the effect of the lumbar epidural catheter used in our study on perioperative cardiac markers was less clearly demonstrated than that of a thoracic epidural catheter. This issue could be clarified by future studies with larger patient groups.

Pro-B-type natriuretic peptide has been widely recognized as a valuable biomarker for the assessment and prediction of cardiac dysfunction in both elective and emergency surgical populations. Elevated perioperative proBNP levels have been associated with increased cardiovascular morbidity and adverse outcomes. In the present study, we aimed to investigate the relationship between epidural analgesia and perioperative proBNP levels in patients undergoing elective vascular surgery and to evaluate its potential impact on cardiac stress and function.

Study Limitations

This study has several limitations that should be considered when interpreting the findings. First, the sample size was relatively small, which may have made it harder to find significant differences in cardiac biomarkers like NT-proBNP and troponin-I between the study groups. Consequently, the lack of statistically significant differences must be regarded with caution, as the study may lack sufficient power to identify minor biomarker variations. Moreover, only early postoperative biomarker levels were analyzed, and long-term clinical cardiac outcomes were not comprehensively evaluated. Additionally, the study cohort comprised solely patients undergoing significant vascular surgery, potentially constraining the applicability of the results to other surgical groups. Advanced cardiac imaging techniques were not utilized, consequently subtle perioperative cardiac alterations may not have been entirely documented. Subsequent investigations involving larger patient populations, diverse patient demographics, and extended postoperative monitoring durations may yield more thorough understanding of the correlation between epidural anesthesia and perioperative cardiac stress. Another limitation of this study is the considerable

age disparity noted between the study groups. Age is known to affect perioperative cardiac biomarkers like NT-proBNP, so this imbalance may have made it harder to understand biomarker results. Future prospective studies with larger sample sizes and more balanced baseline characteristics, incorporating both biomarker assessments and clinically meaningful cardiovascular outcomes, are required to further clarify the potential cardioprotective effects of epidural anesthesia.

Despite these limitations, the prospective randomized design and the comprehensive assessment of hemodynamic parameters, postoperative pain scores, and sequential biomarker measurements establish a hypothesis-generating framework for elucidating the potential impact of epidural anesthesia on the surgical stress response in patients undergoing significant vascular surgery. These findings should therefore be interpreted as hypothesis-generating and warrant confirmation in larger randomized studies evaluating both biomarkers and clinically relevant cardiovascular outcomes.

Conclusion

This study revealed no significant differences between the groups regarding NT-proBNP and troponin-I levels. Our findings suggest that epidural anesthesia during vascular surgery can attenuate increases in (NT-) proBNP and troponin-I and reduce hemodynamic fluctuations. The epidural approach, when applied to the appropriate patients and under safe conditions, may help attenuate changes in blood flow and reduce pain immediately after surgery. Biomarker findings are speculative owing to study constraints and require validation in larger studies that encompass clinical cardiac outcomes.

Ethics

Ethics Committee Approval: The study was conducted in the Cardiovascular Surgery operating room and was approved by the Istanbul University-Cerrahpasa Faculty of Medicine Ethics Committee (approval no.: 43423, date: 01.12.2011).

Informed Consent: Written informed consent was obtained from the responsible adult for each patient.

Footnotes

Authorship Contributions

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

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

Financial Disclosure: The supply of NT-proBNP and troponin I kits was supported by the Scientific Research Projects (BAP) Unit of İstanbul University-Cerrahpaşa Faculty of Medicine. NTproBNP levels were measured using kits from Biomedica (Slovakia), and troponin I levels were measured using kits from Diagnostic Automation Inc.

References

1. Thompson A, Fleischmann KE, SmilowitzNR, et al. AHA/ACC/ACS/ASNC/HRS/SCA/SCCT/SCMR/SVM guideline for perioperative cardiovascular management for noncardiac surgery: a report of the American College of Cardiology/American Heart Association Joint committee on clinical practice guidelines. Circulation. 2024:150:e351-442.
2. Boghean A, Guţu C, Firescu D. Perioperative risk: short review of current approach in non cardiac surgery. J Cardiovasc Dev Dis. 2025;12:24.
3. Halvorsen S, Mehilli J, Cassese S, et al. 2022 ESC Guidelines on cardiovascular assessment and management of patients undergoing non-cardiac surgery. Eur Heart J. 2022;43:3826-924.
4. Jakobsen C-J, Kirkegaard H, Lindskov C, et al. Epidural anaesthesia reduce NT-proBNP after cardiac surgery. Eur J Anaesth. 2002;23:23.
5. Vascular Events In Noncardiac Surgery Patients Cohort Evaluation (VISION) Study Investigators; Devereaux PJ, Chan MT, Alonso-Coello P, et al. Association between postoperative troponin levels and 30-day mortality among patients undergoing noncardiac surgery. JAMA. 2012;307:2295-304.
6. Rodseth RN, Lurati Buse GA, Bolliger D, et al. The predictive ability of pre-operative B-type natriuretic peptide in vascular patients for major adverse cardiac events: an individual patient data meta-analysis. J Am Coll Cardiol. 2011;58:522-9.
7. Loick HM, Schmidt C, Van Aken H, et al. High thoracic epidural anesthesia, but not clonidine, attenuates the perioperative stress response via sympatholysis and reduces the release of troponin T in patients undergoing coronary artery bypass grafting. Anesth Analg. 1999;88:701-9.
8. Vetrugno L, Dogareschi T, Sassanelli R, et al. Thoracic ultrasound evaluation and B-type natriuretic peptide value in elective cesarean section under spinal anesthesia. Ultrasound J. 2020;12:10.
9. Hu Q, Yu X, Zhou T. Epidural analgesia during surgery and its relation to postoperative myocardial infarction: meta-analysis. Wideochir Inne Tech Maloinwazyjne. 2024;19:11-24.
10. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. JAMA. 2013;310:2191-4.
11. Sayed Masri SNN, Basri F, Yunus SN, Cheah SK. Cardiac troponin as a prognostic indicator for major adverse cardiac events in non-cardiac surgery: a narrative review. Diagnostics. 2025; 15:1061.
12. Choi JH, Cho DK, Song YB, et al. Preoperative NT-proBNP and CRP predict perioperative major cardiovascular events in non-cardiac surgery. Heart. 2010;96:56-62.
13. Li Y, Dong H, Tan S, Qian Y, Jin W. Effects of thoracic epidural anesthesia/analgesia on the stress response, pain relief, hospital

stay, and treatment costs of patients with esophageal carcinoma undergoing thoracic surgery: a single-center, randomized controlled trial. Medicine (Baltimore). 2019;98:e14362.

14. Hewson DW, Tedore TR, Hardman JG. Impact of spinal or epidural anaesthesia on perioperative outcomes in adult noncardiac surgery: a narrative review of recent evidence. Br J Anaesth. 2024;133:380-99.
15. Macrosson D, Beebeejaun A, Odor PM. A systematic review and meta-analysis of thoracic epidural analgesia versus other analgesic techniques in patients post-oesophagectomy. Perioper Med (Lond). 2024;13:80.
16. Landoni G, Isella F, Greco M, Zangrillo A, Royse CF. Benefits and risks of epidural analgesia in cardiac surgery. Br J Anaesth. 2015;115:25-32.
17. Clerico A, Del Ry S, Maffei S, Prontera C, Emdin M, Giannessi D. The circulating levels of cardiac natriuretic hormones in healthy adults: effects of age and sex. Clin Chem Lab Med. 2002;40:371-7.
18. Kurt N, Kaşıkçı MM, Pişkin Ö, Hancı V, Sakin M. Comparison of brain natriuretic peptide (BNP) levels in patients under general or spinal anesthesia. Haydarpasa Numune Med J. 2019;59:128-35.
19. Pölzl L, Lohmann R, Theurl F, et al. Prognostic implications of preoperative N-Terminal pro-B-type natriuretic peptide dynamics in patients undergoing cardiac surgery. JACC Adv. 2025;4:102096.
20. Laferrière-Langlois P, Jeffries S, Harutyunyan R, Hemmerling TM. Epidural catheterization in cardiac surgery: a systematic review and risk assessment of epidural hematoma. Ann Card Anaesth. 2024;27:111-20.

Original Article

DOI: 10.4274/haseki.galenos.2026.91885
Med Bull Haseki 2026;64(2):101-107

Prognostic Performance of the PATHOS Score Compared with CURB-65 and A-DROP in Emergency Department Patients with Community-acquired Pneumonia

Kamil Kokulu, Ekrem Taha Sert

Aksaray University Faculty of Medicine, Department of Emergency Medicine, Aksaray, Türkiye

Abstract

Aim: Accurate early risk stratification is essential for guiding disposition and monitoring decisions in community-acquired pneumonia (CAP) in the emergency department. We aimed to evaluate the prognostic performance of the PATHOS score and compare it with CURB-65 and A-DROP in predicting 30-day mortality in adult patients with CAP.

Methods: This retrospective, single-center observational cohort study included consecutive adult patients (≥18 years) presenting to the emergency department with CAP between January 1, 2019, and January 1, 2024. Patient data were obtained from the hospital's electronic medical record system using the International Classification of Diseases, Tenth Revision. CURB-65, A-DROP, and PATHOS scores were calculated based on the collected data. The ability of the PATHOS score to predict 30-day mortality was evaluated and compared with CURB-65 and A-DROP scores.

Results: A total of 605 cases were included in the study. The overall 30-day mortality rate was 8.6%. In predicting 30-day mortality, the PATHOS score [area under the curve (AUC)=0.849] demonstrated better performance than CURB-65 (AUC=0.733) and A-DROP (AUC=0.780). A PATHOS score of >3 was identified as the optimal threshold for predicting 30-day mortality, with a sensitivity of 75.0%, a specificity of 86.8%, and a negative predictive value of 97.4%.

Conclusion: Our findings indicate that the PATHOS score can serve as an effective tool in the clinical management of patients with CAP and contribute significantly to clinical decision-making.

Keywords: Pneumonia, prognosis, risk assessment, clinical decision-making, critical care, community-acquired infections

Introduction

Community-acquired pneumonia (CAP) remains a major cause of morbidity and mortality worldwide despite advances in antimicrobial therapy and supportive care (1). Mortality among hospitalized CAP patients is reported to range from approximately 5% to 15%, rising substantially in older patients and in those with comorbidities or severe disease (2). Contemporary guideline updates continue to emphasize early diagnosis and severity assessment to guide site-of-care decisions and monitoring intensity in adults with CAP (1,3,4).

Several scoring systems are used to estimate disease severity and short-term mortality risk in CAP, including the pneumonia severity index (PSI), CURB-65, and A-DROP (5,6). Although PSI is widely recommended and validated, its multi-variable structure may limit practicality in time-pressured emergency department (ED) settings (5,7). CURB-65 and A-DROP are simpler bedside tools; however, because they rely on a limited set of physiologic and laboratory variables, they may not fully capture risk heterogeneity, particularly among older patients and those with cardiovascular vulnerability (6,7). Accordingly, recent studies continue to develop and evaluate pragmatic

Corresponding Author: Kamil Kokulu, MD, Aksaray University Faculty of Medicine, Department of Emergency Medicine, Aksaray, Türkiye
E-mail: drkokulu@gmail.com **ORCID:** orcid.org/0000-0002-6132-0898
Received: 30.12.2024 **Accepted:** 26.02.2026 **Publication Date:** 27.03.2026

Cite this article as: Kokulu K, Sert ET. Prognostic performance of the PATHOS score compared with CURB-65 and A-DROP in emergency department patients with community-acquired pneumonia. Med Bull Haseki. 2026;64(2):101-107

prediction models and scores for 30-day mortality in CAP (7-9). This ongoing effort reflects the need for risk stratification tools that are both accurate and feasible for application in the ED. In parallel, routinely measured biomarkers such as troponin (reflecting myocardial injury) and platelet count abnormalities have repeatedly been associated with worse outcomes in CAP, suggesting that composite models incorporating these parameters may enhance prognostic discrimination (10,11).

We hypothesized that the PATHOS score comprising age, platelet count, troponin level, heart rate (HR), oxygenation, and systolic blood pressure (SBP) would demonstrate superior prognostic performance for 30-day mortality compared with that of CURB-65 and A-DROP in adult patients presenting to the ED with CAP. Therefore, the aim of this study was to evaluate the ability of PATHOS to predict 30-day mortality and to directly compare its discrimination with that of CURB-65 and A-DROP. We expect that improved early risk stratification at ED presentation will contribute to more consistent disposition decisions (outpatient vs. inpatient care) and to identifying patients who may benefit from closer monitoring and timely escalation of care.

Materials and Methods

Compliance with Ethical Standards

The study protocol was approved by the Aksaray University Health Sciences Scientific Research Ethics Committee (approval number: 2024/021, date: 04.04.2024). The study was conducted in accordance with the principles of the Declaration of Helsinki. Given the retrospective nature of the study, the requirement for informed consent was waived by the committee.

Study Design and Participants

This retrospective, single-center, observational cohort study included patients aged 18 years and older who presented consecutively to the ED and were diagnosed with CAP between January 1, 2019, and January 1, 2024. Patients excluded from the study were those who lacked sufficient clinical data and follow-up information; those under 18 years of age; those exhibiting atypical findings on chest radiography or thoracic computed tomography; those who were not diagnosed with pneumonia; those with a history of hospitalization within the previous two weeks; those with concurrent diagnoses of other severe infections; those diagnosed with acute coronary syndrome; those receiving antibiotic or immunosuppressive therapy; those with aspiration-related pneumonia; and pregnant women. The study flowchart is presented in Figure 1.

Data Collection and Definitions

Patient data were obtained from the hospital's electronic medical records system using the International Classification of Diseases-Tenth Revision. The collected variables included age, gender, vital parameters at presentation (SBP, diastolic blood pressure, respiratory

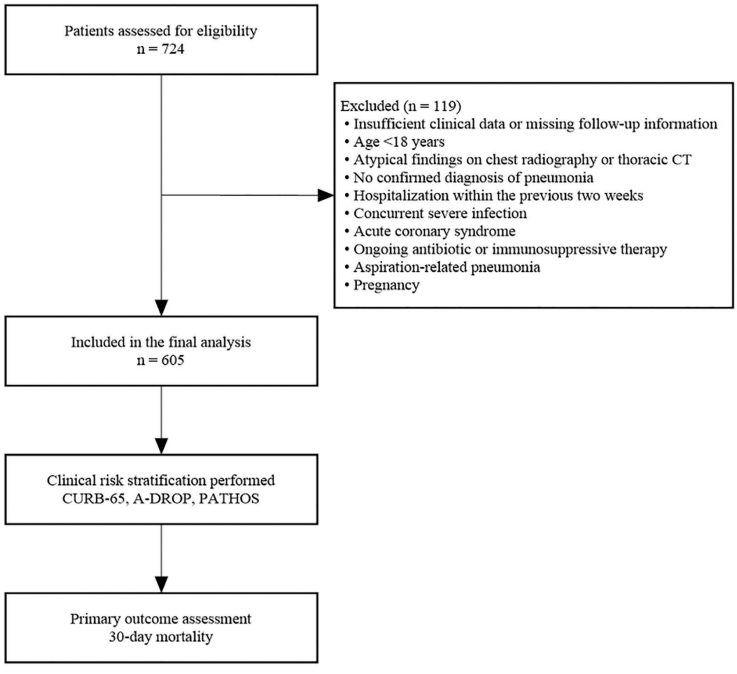

Figure 1. Flow chart of the study
CT: Computed tomography

rate, and HR), complete blood count and biochemical parameters (urea, platelet count, and troponin levels), arterial blood gas parameters [partial arterial oxygen pressure and oxygen saturation (SpO_2)], radiological imaging reports, and patient outcomes (discharge or death). CURB-65, A-DROP, and PATHOS scores were calculated from the obtained data. The PATHOS score assessed the following parameters: age >80 years, platelet count <100 or >400×10³/μL, troponin level >17 pg/mL, SBP<100 mmHg, HR>100 bpm, and SpO_2<90% (12). Each criterion was assigned a score of 1 point if present. Community-acquired pneumonia was defined as the presence of a new infiltrate consistent with pneumonia on chest imaging and at least one of the following clinical findings: fever (≥38 °C) or hypothermia (<35 °C), cough, new or purulent sputum production, dyspnea, pleuritic chest pain, or abnormal breath sounds on auscultation (13). The predictive power of the PATHOS score for 30-day mortality was calculated and compared with CURB-65 and A-DROP scores.

Statistical Analysis

Statistical analyzes were performed using the SPSS, version 22.0 (IBM Corp., Armonk, NY, USA). The normality of the continuous data distribution was evaluated using the Kolmogorov-Smirnov test. Continuous variables with normal distributions were expressed as mean ± standard deviation, while non-normally distributed continuous variables were expressed as median and interquartile range (IQR, 25th-75th percentile). Frequencies and percentages were used for categorical variables. Comparisons of categorical variables were undertaken using the chi-square test. The patients were categorized into two groups based on 30-day mortality outcomes: non-survivors and survivors. Potential risk factors were analyzed using logistic regression, with odds ratio (OR) and 95% confidence intervals (CIs) calculated. Variables with p<0.10 in univariate analysis were entered into the multivariate logistic regression model. Model calibration was assessed using the Hosmer-Lemeshow goodness-of-fit test. The prognostic power of CURB-65, A-DROP, and PATHOS scores in patients with CAP was evaluated using receiver operating characteristic (ROC) analysis. Values for area under the curve (AUC), cutoff, sensitivity, and specificity were determined. The optimal cut-off values were determined using the Youden index from ROC analysis. A p-value of <0.05 was considered statistically significant.

Results

A total of 605 cases were included in the study, comprising 58.8% (n=356) males and 41.2% (n=249) females. The median age of the cases was 68 years (IQR=59-73). The distribution of patients according to their PATHOS, A-DROP, and CURB-65 scores is presented in Table 1.

Subgroup analyses revealed that among patients with PATHOS scores of 0-1, the mortality rate was 2.6% (13/493), compared with 37.6% (32/85) among those with scores of 3-6 (p<0.001). Compared with CURB-65 and A-DROP scores, the PATHOS score provided superior classification of high-risk patients (Table 2).

In logistic regression analysis, significant risk factors for 30-day mortality included age ≥65 years (OR=2.83; 95% CI=1.35-6.17; p=0.011), blood urea nitrogen level >20 mg/dL (OR=2.04; 95% CI=1.37-5.92; p=0.004), platelet count <100×10³/μL or >400×10³/μL (OR=2.69; 95% CI=1.51-4.81; p=0.006), troponin level above the cut-off (OR=3.54; 95% CI=2.18-5.21; p<0.001), SpO_2<90% (OR=1.11; 95% CI=1.04-2.65; p=0.002), and SBP<100 mmHg (OR=1.19; 95% CI=1.07-3.71; p=0.001) (Table 3).

Table 1. Baseline characteristics of patients with community-acquired pneumonia

Variable	
Age, years	68 (59-73)
Gender	
Male	356 (58.8%)
Female	249 (41.2%)
CURB-65 score	
C: confusion	23 (3.8%)
U: blood urea nitrogen >20 mg/dL	129 (21.3%)
R: respiratory rate ≥30 breaths per minute	147 (24.3%)
B: systolic blood pressure <90 mmHg or diastolic ≤60 mmHg	104 (17.2%)
A: age ≥65 years	371 (61.3%)
A-DROP score	
A: age (years) (male ≥70 and female ≥75)	250 (41.3%)
D: dehydration (blood urea nitrogen ≥21 mg/dL)	121 (20.0%)
R: respiratory failure (SaO_2≤90% or PaO_2≤60 mmHg)	75 (12.4%)
O: orientation disruption (confusion)	23 (3.8%)
P: low blood pressure (systolic blood pressure ≤90 mmHg)	104 (17.2%)
PATHOS score	
P: platelet count <100 or >400×10³/μL	63 (10.4%)
A: age>80 years	39 (6.4%)
T: troponin level >cut-off	61 (10.1%)
H: heart rate >100 bpm	132 (21.8%)
O: oxygenation (SpO_2<90%)	75 (12.4%)
S: systolic blood pressure <100 mmHg	123 (20.3%)
30-day mortality	52 (8.6%)
Data are presented as median (25th-75th percentile or n (%)	

The logistic regression model demonstrated good calibration (Hosmer-Lemeshow p>0.05).

The ability of the PATHOS score to predict mortality in patients with CAP was evaluated using ROC analysis (Figure 2). The PATHOS score demonstrated greater accuracy in predicting 30-day mortality (AUC=0.849) than CURB-65 (AUC=0.733) and A-DROP (AUC=0.780).

A PATHOS score of >3 was found to predict 30-day mortality with a sensitivity of 75.0%, a specificity of 86.8%, and a negative predictive value of 97.4% (Table 4).

Discussion

The study results demonstrate that the PATHOS score (AUC=0.849) offers higher accuracy in predicting mortality among patients with CAP compared to CURB-65 (AUC=0.733) and A-DROP (AUC=0.780). In ROC analysis, the PATHOS score predicted mortality with a sensitivity of 75.0% and a specificity of 86.8%. These findings suggest that the PATHOS score may serve as a valuable tool in the clinical management of patients with CAP, addressing gaps in existing scoring systems and contributing to clinical decision-making. Importantly, our study reflects a real-world ED cohort, which may explain the slightly higher mortality rate compared with outpatient CAP populations.

Patients with CAP are generally older individuals with multiple comorbidities (14). Making appropriate treatment decisions is critically important. Clinicians, therefore, require supportive parameters during the decision-making process, and various scoring systems have been developed for this purpose. CURB-65, A-DROP, and PSI are widely used to assess the severity of CAP (5,6). Pneumonia severity index is the recommended scoring system in international guidelines for determining the need for hospitalization, instead of CURB-65 (5).

Pneumonia severity index is based on 20 variables and requires radiographic assessment. These features can complicate its use in crowded EDs and may delay time-sensitive decision-making. Although PSI is widely recommended in guidelines, it was not included in the present analysis due to its complexity and limited availability of some variables in retrospective ED data. Therefore, there is a need for simpler and more rapidly applicable scoring systems in emergency care settings. CURB-65 offers simplicity due to its reliance on a limited number of parameters, enabling quick calculations. However, it may fall short of accurately determining the severity of pneumonia in patients with comorbidities.

Previous studies have reported CURB-65 AUC values of 0.829 for 28-day mortality (15), 0.755 for 30-day mortality (16), and 0.738 for in-hospital mortality (17).

Table 2. Thirty-day mortality rates across subgroups of the CURB-65, A-DROP and PATHOS scores

Subgroup	All	30-day mortality
CURB-65 score		
0-1	390 (64.5%)	14 (3.6%)
2	137 (22.6%)	18 (13.1%)
3-5	78 (12.9%)	20 (25.6%)
p-value		<0.001
A-DROP score		
0-1	458 (75.7%)	14 (3.1%)
2	88 (14.5%)	23 (26.1%)
3-5	59 (9.8%)	15 (25.4%)
p-value		<0.001
PATHOS score		
0-1	493 (81.5%)	13 (2.6%)
2	27 (4.5%)	7 (25.9%)
3-6	85 (14.0%)	32 (37.6%)
p-value		<0.001

Statistical tests: Fisher's exact test for categorical variables

Table 3. Univariate and multivariate analyses of risk factors associated with 30-day mortality in patients with community-acquired pneumonia

	Univariate logistic regression			Multivariate logistic regression		
Risk factors	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Age ≥65 years	2.33	1.31-4.15	0.004	2.83	1.35-6.17	0.011
Confusion	1.89	0.23-4.43	0.989	1.71	0.29-8.09	0.549
Respiratory rate ≥30/min	1.74	0.95-3.18	0.072	1.53	0.52-4.68	0.452
Heart rate >100 bpm	1.51	1.06-2.34	0.048	1.22	0.46-1.89	0.114
Systolic blood pressure <100 mmHg	1.75	1.16-2.34	<0.001	1.19	1.07-3.71	0.001
SpO_2<90%	1.21	1.09-1.95	<0.001	1.11	1.04-2.65	0.002
Blood urea nitrogen >20 mg/dL	3.38	2.17-7.46	<0.001	2.04	1.37-5.92	0.004
Platelet count <100 or >400×10^3/μL	3.75	2.49-5.39	<0.001	2.69	1.51-4.81	0.006
Troponin level >cut-off	4.96	3.10-7.31	<0.001	3.54	2.18-5.21	<0.001

CI: Confidence interval, OR: Odds ratio

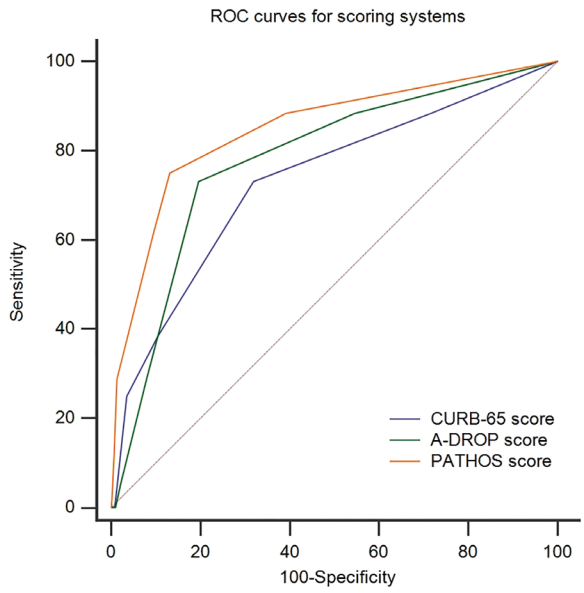

Figure 2. Receiver operating characteristic curves for the scoring systems
ROC: Receiver operating characteristic

Studies have also shown that A-DROP performs similarly to CURB-65 in evaluating disease severity (6,18). A recent study comparing 30-day mortality rates found comparable discriminative performance between A-DROP and CURB-65 (AUC=0.756 for A-DROP and 0.808 for CURB-65). However, both scores demonstrated a lower ability to predict 72-hour ED revisit rates (AUC=0.617 for A-DROP and 0.639 for CURB-65) (6). Variability in reported AUC values complicates achieving consensus regarding these systems' effectiveness. Our findings are consistent with previous studies regarding the performance of CURB-65 and A-DROP scores in predicting mortality.

In the literature, differences in patient selection and event rates appear to influence the performance of bedside scores. In a prospective Japanese cohort of hospitalized CAP patients, Kasamatsu et al. (18) reported high discrimination for A-DROP and CURB-65 (AUC=0.88 for each) and for PSI (AUC=0.89). In the same cohort, CRP performed poorly (AUC=0.54), whereas semi-quantitative procalcitonin showed intermediate discrimination (AUC=0.80) (18). By contrast, Limapichat and Supavajana (6) evaluated an ED cohort largely considered for discharge with a very low 30-day mortality rate (1.47%) and reported AUCs of 0.756 for A-DROP and 0.808 for CURB-65, alongside only modest discrimination for 72-hour ED revisit (AUC=0.617 vs. 0.639, respectively) (6). In our ED cohort, which had a higher 30-day mortality rate (8.6%) and a broader spectrum of severity, A-DROP (AUC=0.780) and CURB-65 (AUC=0.733) demonstrated performance within the ranges reported in prior studies. In contrast, the PATHOS score demonstrated greater discriminatory ability, with an AUC of 0.849. These comparisons suggest that cohort characteristics may partly explain heterogeneous A-DROP and CURB-65 performance and also support the potential added value of incorporating troponin and platelet abnormalities in ED-based mortality risk stratification.

Previous research has highlighted the prognostic significance of troponin and platelet levels in patients with CAP. Platelets play a vital role in antimicrobial defense and modulation of inflammatory responses. Studies on platelet count abnormalities (thrombocytopenia and thrombocytosis) in patients with CAP have demonstrated an association between these abnormalities and clinical outcomes. Mirsaeidi et al. (10) reported that both thrombocytopenia and thrombocytosis were associated with 30-day mortality, with thrombocytosis carrying a higher mortality risk. In another study, Ghoneim et al. (19) identified associations between thrombocytosis and respiratory complications such as empyema, lung abscess, and pleural effusion. Thrombocytopenia was linked to sepsis and septic shock. Both abnormalities independently predicted 30-day mortality. In contrast, Mısıroğlu et al. (20) found that platelet changes were not directly predictive of mortality but were associated with increased intensive care and mechanical ventilation needs during follow-up. These findings suggest that platelet count should be considered a significant biomarker for predicting mortality and complications in patients with CAP.

Troponin is a biomarker frequently used to assess myocardial injury. Studies have shown that troponin levels measured at the time of presentation are associated with increased mortality in patients with CAP. A study involving 295 patients with CAP found that elevated troponin levels at presentation were a strong predictor of both short- and long-term mortality risks (11). The same study demonstrated that combining troponin levels with the

Table 4. The accuracy of different scoring systems in predicting 30-day mortality

Scoring system	AUC (95% CI)	Cut-off	Sensitivity	Specificity	PPV	NPV
CURB-65 score	0.733 (0.695-0.767)	>3	73.06	67.99	17.7	96.4
A-DROP score	0.780 (0.744-0.812)	>3	73.08	80.29	25.9	96.9
PATHOS score	0.849 (0.818-0.877)	>3	75.00	86.80	34.8	97.4

AUC: Area under the curve, CI: Confidence interval, PPV: Positive predictive value, NPV: Negative predictive value

PSI classification provided superior predictive capacity for short-term mortality compared with the PSI classification alone (AUC for troponin and PSI combination=0.903; AUC for PSI alone=0.818). Similarly, a study focusing on patients with CAP without acute coronary syndrome identified an association between elevated troponin levels and mortality in the intensive care unit (21). The increase in troponin levels is considered related to oxygen supply-demand imbalance, systemic inflammation, and coagulation mechanisms. Our study highlights the effectiveness of the PATHOS score as a tool for predicting prognosis in patients with CAP. This score incorporates platelet count and troponin levels, thereby improving its accuracy and reliability. The inclusion of additional parameters, such as troponin and platelet count, in scoring systems can enhance the precision of risk stratification, particularly for patients with high cardiovascular risk. Therefore, the PATHOS score is a robust alternative with a broader set of parameters for determining prognosis in this patient population. In addition, PATHOS has shown prognostic utility in other ED-based acute care cohorts, supporting further evaluation across ED populations (22,23).

Study Limitations

This study has certain limitations. First, it had a retrospective design. Second, it was conducted at a single center, which may have limited sample size and patient diversity, thereby reducing the generalizability of the findings to other patient groups. Third, scores were calculated from baseline presentation variables and therefore do not account for dynamic changes during treatment; consequently, our findings primarily inform early risk stratification rather than assessment of response to therapy. In addition, the predictive efficacy of the PATHOS score lacked external validation in an independent cohort, necessitating cautious interpretation of the findings' generalizability.

Despite these limitations, our study provides a consecutive, ED-based adult CAP cohort with short-term outcome assessment and a direct, clinically interpretable head-to-head comparison of three bedside tools using both discrimination metrics and risk-stratum separation.

Conclusion

In this ED-based cohort of adults with CAP, PATHOS demonstrated greater discriminative ability for predicting 30-day mortality than CURB-65 and A-DROP. With a cut-off of >3, PATHOS provided high specificity and an excellent negative predictive value, supporting its use as a practical bedside option for early risk stratification in the ED.

Ethics

Ethics Committee Approval: The study protocol was approved by the Aksaray University Health Sciences Scientific Research Ethics Committee (approval number: 2024/021, date: 04.04.2024).

Informed Consent: Given the retrospective nature of the study, the requirement for informed consent was waived by the committee. All patient data were anonymized prior to analysis.

Footnotes

Authorship Contributions

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

Conflict of Interest: No conflicts of interest or competing interests have been reported by the authors or any individuals with control over the content of this article.

Financial Disclosure: No financial support was received for this paper or for the research that resulted in it.

References

1. Jones BE, Ramirez JA, Oren E, et al. Diagnosis and management of community-acquired pneumonia: an Official American Thoracic Society Clinical Practice Guideline. Am J Respir Crit Care Med. 2026;212:24-44.
2. Ferreira-Coimbra J, Sarda C, Rello J. Burden of community-acquired pneumonia and unmet clinical needs. Adv Ther. 2020;37:1302-18.
3. Dinh A, Barbier F, Bedos JP, et al. Update of guidelines for management of Community Acquired pneumonia in adults by the French Infectious Disease Society (SPILF) and the French-Speaking Society of Respiratory Diseases (SPLF): Endorsed by the French intensive care society (SRLF), the French microbiology society (SFM), the French radiology society (SFR) and the French emergency society (SFMU). Respir Med Res. 2025;87:101161.
4. National Institute for Health and Care Excellence (NICE). Pneumonia: diagnosis and management [Internet]. London: NICE; 2025 Sep 2 [cited 2026 Jan 28]. Available from: https://www.nice.org.uk/guidance/ng250
5. Metlay JP, Waterer GW, Long AC, et al. Diagnosis and treatment of adults with community-acquired pneumonia. An official clinical practice guideline of the American Thoracic Society and Infectious Diseases Society of America. Am J Respir Crit Care Med. 2019;200:e45-67.
6. Limapichat T, Supavajana S. Comparison between the severity scoring systems A-DROP and CURB-65 for predicting safe discharge from the emergency department in patients with community-acquired pneumonia. Emerg Med Int. 2022;2022:6391141.
7. Pomerantz A, Yaacov A, Goldman A, et al. Evaluation of PSI, CURB-65, and CRB-65 as prognostic tools in hospitalized

immunocompetent patients with pneumonia: real-world outcomes from internal medicine wards. Pneumonia (Nathan). 2026;18:3.

8. Algin A, Özdemir S, Afacan MA, Yusufoglu K, Özkan A. Comparative analysis of APUA, APUA-RO2, and CURB-65 scores for mortality risk in hospitalized pneumonia patients. Am J Emerg Med. 2025;97:136-139.
9. Calis AG, Karaboga B, Uzer F, et al. Correlation of pneumonia severity ındex and CURB-65 score with neutrophil/lymphocyte ratio, platelet/lymphocyte ratio, and monocyte/lymphocyte ratio in predicting ın-hospital mortality for community-acquired pneumonia: observational study. J Clin Med. 2025;14:728.
10. Mirsaeidi M, Peyrani P, Aliberti S, et al. Thrombocytopenia and thrombocytosis at time of hospitalization predict mortality in patients with community-acquired pneumonia. Chest. 2010;137:416-420.
11. Vestjens SMT, Spoorenberg SMC, Rijkers GT, et al. High-sensitivity cardiac troponin T predicts mortality after hospitalization for community-acquired pneumonia. Respirology. 2017;22:1000-1006.
12. Spampinato MD, Covino M, Passaro A, et al. Predicting in-hospital mortality in pulmonary embolism patients: development and external validation of the PATHOS score. Clin Exp Emerg Med. 2023;10:26-36.
13. Carratalà J, Fernández-Sabé N, Ortega L, et al. Outpatient care compared with hospitalization for community-acquired pneumonia: a randomized trial in low-risk patients. Ann Intern Med. 2005;142:165-172.
14. Putot A, Tetu J, Perrin S, et al. A new prognosis score to predict mortality after acute pneumonia in very elderly patients. J Am Med Dir Assoc. 2016;17:1123-1128.
15. Zhou H, Guo S, Lan T, Ma S, Zhang F, Zhao Z. Risk stratification and prediction value of procalcitonin and clinical severity scores for community-acquired pneumonia in ED. Am J Emerg Med. 2018;36:2155-2160.
16. Ito A, Ishida T, Tokumasu H, Yamazaki A, Washio Y. Evaluation of pneumonia severity scoring systems in nursing and healthcare-associated pneumonia for predicting prognosis: a prospective, cohort study. J Infect Chemother. 2020;26:372-378.
17. Ma C-M, Wang N, Su Q-W, Yan Y, Yin F-Z. The performance of CURB-65 and PSI for predicting in-hospital mortality of community-acquired pneumonia in patients with type 2 diabetes compared with the non-diabetic population. Diabetes Metab Syndr Obes. 2021;14:1359-1366.
18. Kasamatsu Y, Yamaguchi T, Kawaguchi T, et al. Usefulness of a semi-quantitative procalcitonin test and the A-DROP Japanese prognostic scale for predicting mortality among adults hospitalized with community-acquired pneumonia. Respirology. 2012;17:330-336.
19. Ghoneim AHA, Mohammad MA, Elghamrawy MA, Embarak S. Platelet count as a predictor of outcome of hospitalized patients with community-acquired pneumonia at Zagazig University Hospitals, Egypt. Egypt J Bronchol. 2020;14:11.
20. Mısıroğlu RS, Aksay E, Şancı E, Tertemiz KC. Can thrombocytosis or thrombocytopenia predict complicated clinical course and 30-days mortality in patients with pneumonia? Turk J Med Sci. 2021;51:2903-2907.
21. Lee YJ, Lee H, Park JS, et al. Cardiac troponin I as a prognostic factor in critically ill pneumonia patients in the absence of acute coronary syndrome. J Crit Care. 2015;30:390-394.
22. Menekse TS, Gunsay RH, Sert ET, Utlu SG, Kokulu K. PATHOS score as a predictor of in-hospital mortality in patients with acute cardiogenic pulmonary edema presenting to the emergency department. Gazi Med J. 2025;36:422-429.
23. Sert ET, Kokulu K. Performance of the PATHOS score in predicting in-hospital mortality in patients aged 65 years and older admitted to the intensive care unit from the emergency department. Cukurova Med J. 2025;50:99-105.

Original Article

DOI: 10.4274/haseki.galenos.2026.55265
Med Bull Haseki 2026;64(2):108-115

The Relationship Between Mindful Parenting and Parental Burnout in Mothers of Children with Attention-deficit Hyperactivity Disorder

Orhan Kocaman[1], Merve Sema Arslan[1], Tacettin Kuru[2], Tayfun Kara[1], Pinar Aydogan Avsar[3]

[1]*Alanya Alaaddin Keykubat University Faculty of Medicine, Department of Child and Adolescent Psychiatry, Antalya, Türkiye*

[2]*Alanya Education and Research Hospital, Clinic of Psychiatry, Antalya, Türkiye*

[3]*Alanya Education and Research Hospital, Clinic of Child and Adolescent Psychiatry, Antalya, Türkiye*

Abstract

Aim: Parental burnout is common among mothers of children with attention deficit hyperactivity disorder (ADHD); however, evidence regarding protective psychological factors is limited. This study aimed to examine mindful parenting and child symptom severity as determinants of burnout among mothers of children with ADHD.

Methods: This cross-sectional study included mothers of 127 children diagnosed with ADHD. The Mindfulness in Parenting Questionnaire (MIPQ), Maslach Burnout Inventory (MBI), and Children and Adolescent Behavior Inventory were administered to the participants. Associations were examined using correlation analyses, followed by multiple linear regression models to identify the determinants of maternal burnout dimensions.

Results: Mindfulness in Parenting Questionnaire-parental self-efficacy (MIPQ-PSE) and being in the moment with the child (MIPQ-BMC) were negatively correlated with depersonalization (MBI-DP) and emotional exhaustion (MBI-EE) and positively correlated with personal accomplishment (MBI-PA) ($p<0.05$). Hyperactivity-impulsivity and attention deficit were positively correlated with MBI-EE and MBI-DP, and negatively correlated with MBI-PA ($p<0.05$). Regression analysis showed that MIPQ-PSE negatively predicted MBI-EE and MBI-DP, whereas hyperactivity-impulsivity positively predicted them ($p<0.05$). Maslach Burnout Inventory-personal accomplishment was positively predicted by MIPQ-PSE and MIPQ-BMC and negatively predicted by attention deficit ($p<0.05$).

Conclusion: Mindful parenting is associated with reduced maternal burnout. Interventions that promote mindfulness have the potential to mitigate burnout among mothers of children with ADHD.

Keywords: Attention deficit-hyperactivity disorder, mindfulness, parenting, burnout

Introduction

Attention deficit hyperactivity disorder (ADHD) is a neurodevelopmental condition that begins in childhood and is characterized by a decreased attention span, impulsivity, and hyperactivity (1). The global prevalence in children and adolescents ranges between 3% and 7.6% (2,3), and it is one of the most common childhood psychiatric disorders (4). Children with ADHD tend to behave less collaboratively and exhibit more negative behavior than their peers. This causes parents to face greater difficulty in coping with damaging behaviors that emerge in such settings as school, the home, or the social environment (5). A child with ADHD behaving in a manner such as to require constant observation and intervention can gradually lead to the development of burnout symptoms by depleting the parent's psychological

Corresponding Author: Assoc. Prof., Tacettin Kuru, Alanya Education and Research Hospital, Clinic of Psychiatry, Antalya, Türkiye
E-mail: Tkuru76@yahoo.com **ORCID:** orcid.org/0000-0002-0313-558X
Received: 11.06.2025 **Accepted:** 03.03.2026 **Publication Date:** 27.03.2026

Cite this article as: Kocaman O, Arslan MS, Kuru T, Kara T, Aydogan Avsar P. the relationship between mindful parenting and parental burnout in mothers of children with attention-deficit hyperactivity disorder. Med Bull Haseki. 2026;64(2):108-115

resources (6). Studies have shown that symptoms such as chronic stress, mental problems, and burnout are more common in the parents, particularly mothers, of children diagnosed with ADHD (6,7). However, existing research has predominantly emphasized child-related risk factors, while psychological resources that may buffer parental burnout remain unexplored. In recent years, the concept of mindfulness in parenting has emerged as a potential protective psychological resource (8,9).

Mindfulness in parenting involves responding to the child's behavior with awareness rather than an emotional reaction, understanding the child's need in an empathetic manner, and being a more conscious entity in parent-child interactions (10). Mindfulness in parenting can reduce adverse experiences such as impatience, hopelessness, and burnout by increasing parental emotional regulation (11). Research shows that a high degree of mindfulness can help parents develop more effective strategies for coping with their children's difficult behavior and that they will experience less parental stress and will be able to establish more positive parent-child relationships (12,13). From that perspective, a high level of mindfulness in parenting can also be regarded as an important protective factor against parental burnout (14). Although the effects of ADHD symptom severity on parental burnout have been examined within the framework of risk factors, the combined contribution of mindful parenting and child symptom severity to maternal burnout has not been adequately investigated. We hypothesized that higher levels of mindful parenting are associated with lower maternal burnout, whereas greater severity of child ADHD symptoms is associated with higher maternal burnout.

The present study aimed to examine mindful parenting and the severity of child symptoms as determinants of burnout in mothers of children with ADHD. By addressing both risk and protective factors, this study aims to provide clinically relevant evidence to inform supportive interventions for mothers of children with ADHD.

Materials and Methods

Compliance With Ethical Standards

The study was completed with 127 participants. Approval for the study was granted by the Alanya Alaaddin Keykubat University Faculty of Medicine Clinical Research Ethics Committee (approval no.: 02-08, date: 22.01.2025). This study was conducted in accordance with the principles of the Declaration of Helsinki. The purpose of the study was explained to the participants before commencement, and written and verbal consent was obtained from all participants.

Study Design and Population

This was a cross-sectional study. The participants were mothers of children aged 6-18 who presented to the Alanya Education and Research Hospital Child and Adolescent Psychiatry Clinic, Türkiye, and who were diagnosed with ADHD. Participants were consecutively recruited according to their order of presentation at the clinic. The sample size was determined based on the number of eligible participants who attended the clinic during the study period. The study included 149 mothers of children diagnosed with ADHD. All children included in the study had a confirmed diagnosis of attention-deficit/hyperactivity disorder and were receiving ongoing ADHD treatment at the time of assessment. None of the children had been newly diagnosed. All children underwent psychiatric evaluation based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition, Text Revision. Comorbid psychiatric and chronic diseases in children were adopted as exclusion criteria since these might create independent effects on maternal burnout. Twenty-two children with comorbid psychiatric and chronic disorders were excluded. Children with chronic diseases such as neurological, genetic, or metabolic disorders and their siblings were also included.

Psychometric Instruments

The study data were collected through face-to-face interviews with mothers at the time of presentation to the clinic. The mothers of children/adolescents diagnosed with ADHD were administered a sociodemographic data form, the Mindfulness in Parenting Questionnaire (MIPQ), and the Maslach Burnout Inventory (MBI). The Children and Adolescent Behavior Inventory (CABI)-parent version was used to assess the severity of children's ADHD.

Mindfulness in Parenting Questionnaire: This 28-item Likert-type scale was developed by McCaffrey et al. (15) to evaluate mindfulness in parent-child interaction. It was adapted into Turkish by Gördesli et al. (16), resulting in 24 items and two subdimensions, parental self-efficacy (MIPQ-PSE) and being in the moment with the child (MIPQ-BMC), which were obtained in the validity and reliability study. The MIPQ has no cut-off point. Mindfulness levels were evaluated based on the subdimensions. Higher scores indicate higher levels of mindfulness.

Maslach Burnout Inventory: The MBI is a 22-item Likert-type scale developed by Maslach and Jackson (17) to measure individuals' burnout levels. The validity and reliability of the MBI were confirmed by Ergin (18). It consists of three subscales: emotional exhaustion (MBI-EE), depersonalization (MBI-DP), and personal accomplishment (MBI-PA). The EE and DP subscales contain negative responses, whereas the PA subscale contains positive responses. The MBI has no cut-off point.

Children and Adolescent Behavior Inventory-Parent Version: This scale was developed by Burns et al. (19). The validity and reliability study for Turkish was performed by Çiftçi et al. (20). CABI consists of 67 items and nine subscales: sluggish cognitive tempo, anxiety disorder, depressive disorder, attention deficit, hyperactivity-impulsivity, oppositional-defiant disorder, emotionless personality traits, social disorder, and academic impairment. The CABI is a continuous scale with no cutoff points. The total score was the sum of the subscale item scores. Higher total scores indicate greater symptom severity. Attention-deficit and hyperactivity-impulsivity subscales were employed in the present study.

Statistical Analysis

Statictical analyzes were performed using the Jamovi, version 26.0 (Jamovi Project, Sydney, Australia). Participants' sociodemographic data were summarized using descriptive statistics (frequencies and percentages). Skewness, kurtosis, and graphical methods (histogram and Q-Q plot) were used together to evaluate the distribution characteristics of the scale scores. Skewness and kurtosis values between -1 and +1 were regarded as consistent with normal distribution (21). The internal consistency of the scales was evaluated using Cronbach's alpha. Cronbach's alpha values of 0.70 and above indicate an acceptable level of internal consistency. Pearson's correlation analysis (for normally distributed data) and Spearman's rank correlation coefficient (for non-normally distributed data) were used to examine correlations between variables. The independent sample t-test was applied to compare paired groups exhibiting a normal distribution, and one-way analysis of variance was used for multiple groups. In the case of non-normal distribution, the Mann-Whitney U test was applied for pairwise group comparisons, and the Kruskal-Wallis H test was used for comparisons among multiple groups. The MBI-EE, MBI-DP, and MBI-PA subscales were used as the dependent variables for the primary aim of the research. Multiple linear regression analyses were applied separately to each of these three subscales. For each analysis, child-related variables (gender, age, duration of treatment, level of attention deficit, and level of hyperactivity-impulsivity) and parent-related variables (MIPQ subscales) were included in the model as predictors. The model hypotheses were tested using regression analyses. The normality of the distribution of errors was examined using histograms and P-P plots. The regression scatter plot and Durbin-Watson values (1.5-2.5) were examined to assess the assumption of homoscedasticity. Multicollinearity was assessed using the variance inflation factor (VIF) (VIF <10). Pre-test examination revealed that these assumptions were not violated. P-values <0.05 were regarded as significant for all analyses.

Results

The study included mothers of 127 children with ADHD. The majority of the children were male (73.2%), and most were firstborn (63.0%). Nearly half (47.2%) of the families had two children. Approximately one-third of the mothers had primary education (33.9%), high school education (33.1%), or university-level education (33.1%). More than half of the mothers were unemployed (52.8%), and most families had a nuclear family structure (84.3%). Burnout subscale scores (EE, DP, and PA) did not differ significantly across sociodemographic characteristics (all p>0.05).

Correlation analyses revealed significant positive associations between attention deficit levels and both EE (r=0.409, p<0.001) and DP (r=0.407, p<0.001). Similarly, hyperactivity-impulsivity symptoms were positively correlated with EE (r=0.442, p<0.001) and DP (r=0.383, p<0.001) and negatively correlated with PA (r=-0.192, p=0.030) (Table 1). Significant correlations were also observed between parental mindfulness and burnout. The MIPQ-PSE subdimension was negatively correlated with EE (r=-0.319, p<0.001) and DP (r=-0.352, p<0.001) and positively correlated with PAPA (r=0.426, p<0.001). Similarly, the MIPQ-BMC subdimension exhibited a significant positive correlation with PA (r=0.455, p<0.001) (Table 1).

In the multiple linear regression analysis performed to identify variables predicting maternal burnout levels, the ages of the child and the mother, the sex of the child, duration of treatment, attention deficit, hyperactivity-impulsivity, PSE, and BMC were included in the study. The model in which EE (MBI-EE) was the dependent variable was significant [F(8,118)=7.52, RÇ=0.293, p<0.001]. Hyperactivity-impulsivity levels emerged as a significant and positive predictor of EE (β=0.369, p<0.001). The MIPQ-PSE was a significant negative predictor of EE (β=-0.347, p<0.001) (Table 2).

Regression analysis of the model in which the MBI-DP subscale was the dependent variable was successful [F(8,118)=7.31, RÇ=0.286, p<0.001]. Hyperactivity-impulsivity was a significant positive predictor of the DP burnout subscale (β=0.331, p=0.002). The MIPQ-PSE was a significant negative predictor of DP (β=-0.335, p=0.001) (Table 2).

The final model, in which MBI-PA was the dependent variable, was also significant [F(8,118)=7.46, RÇ=0.291, p<0.001]. Attention-deficit levels had a significant negative effect on PA (β=-0.240, p=0.019). The MIPQ-PSE and MIPQ-BMC significantly and positively predicted PA (β=0.200, p=0.046, and β=0.343, p<0.001, respectively) (Table 2).

Table 1. Analysis of the measurement tool data and the relationships with burnout sub-dimensions

	Mean±SD median (Q1-Q3)	Skew	Kurt	α	MBI-EE	MBI-DP	MBI-PA
					r (p-value)	r (p-value)	r (p-value)
Child's age	10.18±2.74	0.533	-0.447	-	-0.069 (0.443)	0.037 (0.679)	0.006 (0.946)
Mother's age	38.36±6.28	0.300	-0.303	-	-0.076 (0.393)	0.008 (0.926)	-0.009 (0.921)
Duration of treatment	12 (6-30)	1.717	2.765	-	-0.018[s] (0.838)	-0.048[s] (0.590)	0.069[s] (0.443)
Attention deficit	19.76±10.27	0.319	-0.684	0.909	0.409 (**<0.001**)	0.407 (**<0.001**)	-0.320 (**<0.001**)
Hyperactivity-impulsivity	21.83±12.48	0.209	-0.984	0.931	0.442 (**<0.001**)	0.383 (**<0.001**)	-0.192 (**0.030**)
MIPQ-PSE	42.69±5.97	-0.329	-0.626	0.857	-0.319 (**<0.001**)	-0.352 (**<0.001**)	0.426 (**<0.001**)
MIPQ-BMC	36.02±4.58	-0.277	-0.574	0.747	-0.160 (0.073)	-0.219 (**0.014**)	0.455 (**<0.001**)
MBI-EE	9.45±6.28	0.831	0.392	0.815	-	-	-
MBI-DP	5.20±4.33	0.904	0.549	0.793	-	-	-
MBI-PA	24.97±4.64	-0.515	-0.405	0.763	-	-	-

Pearson correlation coefficients were calculated for normally distributed variables, and Spearman rank correlation coefficients were used for non-normally distributed variables. Cronbach's alpha (α) values indicate internal consistency. Statistically significant correlations (p<0.05) are shown in bold
SD: Standard deviation, Skew: Skewness, Kurt: Kurtosis, MBI-EE: Maslach Burnout Inventory-emotional exhaustion, MBI-DP: Maslach Burnout Inventory-depersonalization, MBI-PA: Maslach Burnout Inventory-personal accomplishment, MIPQ-BMC: Mindfulness in Parenting Questionnaire-being in the moment with the child, MIPQ-PSE: Mindfulness in Parenting Questionnaire-parental self-efficacy

Table 2. Regression analysis of factors affecting burnout subscales

				95% CI (for standardized estimate)			
DV	**Predictor**	**Estimate**	**Standardized estimate**	**Lower**	**Upper**	**t**	**p**
MBI-EE	**Child's age**	0.186	0.081	-0.142	0.304	0.721	0.472
	Mother's age	-0.113	-0.112	-0.296	0.071	-1.213	0.228
	Sex (Ref: female)	0.046	0.007	-0.333	0.348	0.042	0.966
	Duration of treatment	-0.003	-0.011	-0.198	0.176	-0.116	0.908
	Attention deficit	0.094	0.154	-0.045	0.354	1.534	0.128
	Hyperactivity-impulsivity	0.186	0.369	0.160	0.579	3.486	**<0.001**
	MIPQ-PSE	-0.365	-0.347	-0.543	-0.151	-3.506	**<0.001**
	MIPQ-BMC	0.024	0.017	-0.182	0.216	0.172	0.864
		F(8-118)=7.52. p<0.001. Adjusted R^2=0.293					
MBI-DP	**Child's age**	0.295	0.187	-0.037	0.411	1.655	0.101
	Mother's age	-0.045	-0.065	-0.249	0.120	-0.695	0.488
	Sex (Ref: female)	0.249	0.058	-0.285	0.400	0.334	0.739
	Duration of treatment	-0.016	-0.085	-0.273	0.103	-0.898	0.371
	Attention deficit	0.076	0.179	-0.021	0.380	1.772	0.079
	Hyperactivity-impulsivity	0.115	0.331	0.120	0.542	3.111	**0.002**
	MIPQ-PSE	-0.243	-0.335	-0.532	-0.138	-3.372	**0.001**
	MIPQ-BMC	-0.018	-0.019	-0.219	0.181	-0.187	0.852
		F(8-118)=7.31. p<0.001. Adjusted R^2=0.286					
MBI-PA	**Child's Age**	0.068	0.040	-0.183	0.263	0.355	0.723
	Mother's age	0.013	0.017	-0.167	0.201	0.185	0.853
	Sex (ref: female)	-0.127	-0.028	-0.369	0.314	-0.159	0.874
	Duration of treatment	0.005	0.026	-0.162	0.213	0.270	0.788
	Attention deficit	-0.108	-0.240	-0.440	-0.041	-2.382	**0.019**
	Hyperactivity-impulsivity	-0.032	-0.085	-0.295	0.125	-0.799	0.426
	MIPQ-PSE	0.155	0.200	0.004	0.396	2.018	**0.046**
	MIPQ-BMC	0.347	0.343	0.144	0.542	3.412	**<0.001**
		F(8-118)=7.46. p<0.001. Adjusted R^2=0.291					

Multiple linear regression analysis was performed using the enter method. Standardized estimates (β) with 95% confidence intervals are presented. Model fit is indicated by F statistics and adjusted R^2 values. Statistically significant results (p<0.05) are indicated in bold font
DV: Dependent variable, 95% CI: 95% Confidence interval, MBI-EE: Maslach Burnout Inventory-emotional exhaustion, MBI-DP: Maslach Burnout Inventory-depersonalization, MBI-PA: Maslach Burnout Inventory-personal accomplishment, MIPQ-BMC: Mindfulness in Parenting Questionnaire-being in the moment with the child, MIPQ-PSE: Mindfulness in Parenting Questionnaire-parental self-efficacy

Discussion

This study examined the effects of mindful parenting and child ADHD severity on burnout among mothers of children with ADHD. Previous studies investigating the relationship between mindfulness and burnout in parents are few. To the best of our knowledge, this study is the first to examine the relationship between mindfulness and burnout levels in the mothers of children with ADHD. The study findings show associations among mothers' mindfulness-in-parenting subscales, ADHD symptom severity levels, and burnout levels. In addition, regression analysis showed that higher mindfulness-in-parenting subscale scores and lower attention-deficit symptoms predicted increased PA among mothers. Decreased PSE and increased child hyperactivity-impulsivity predicted EE and DP in mothers. The findings thus support our first hypothesis that maternal burnout levels will decline as mindfulness in parenting increases. They also support our second thesis that maternal burnout levels increase with the severity of ADHD symptoms in children. This study makes a significant contribution to the literature by enhancing our understanding of the factors that affect maternal burnout.

Although several studies have reported that burnout is frequently observed in the mothers of children with ADHD (6,22), those examining the relationship between parental factors and parental burnout are notably limited in clinical ADHD samples. Studies have suggested that imbalances between risk factors and protective factors during the care of children with chronic and mental diseases may lead to the development of burnout among parents (23). This study found that PSE, a sub-dimension of awareness in parenting, negatively predicted EE and DP (sub-dimensions of burnout) and positively predicted PA. This finding aligns with the findings of recent studies examining the relationship between PSE and overall parental burnout in non-clinical samples (14,24,25) and contributes to the existing literature by providing a detailed analysis of the effect of self-efficacy on burnout sub-dimensions. Parental self-efficacy refers to a parent's self-efficacy in the parent-child relationship (26). Parents' beliefs about their competence reported directly influence their vulnerability to burnout (24). In this context, elevated PSE may act as a significant protective factor, enhancing parents' perceived control and coping skills in the face of persistent ADHD-related challenges, thereby reducing vulnerability to burnout and constituting a clinically important target for preventive interventions.

In contrast to the findings related to PSE, the "BMC" sub-dimension of mindful parenting exhibited a more selective pattern of association with burnout. Despite the correlation between being present with the child and DP, this association was no longer significant in the regression model after controlling for PSE and other variables. This suggests that the effect of being present with the child on DP may be indirect or secondary. Conversely, "BMC" emerged as a robust and independent correlate of PA in both correlation and regression analyses. This pattern suggests that increased mindfulness in parenting during mother-child interactions primarily contributes to the development of positive and rewarding aspects of parenting rather than to the reduction of the negative components of burnout. A preceding study reported a correlation between mindfulness in parenting and mother-child interactions, with parents demonstrating higher levels of mindfulness in their parenting (27). Mothers who demonstrate effective parent-child interaction have been shown to exhibit a higher level of competence in their parenting roles and are thought to experience a greater sense of accomplishment in this role (28). From the perspective of ADHD, it is reported that parents of children with ADHD experience more strained and less satisfying interactions with their children compared to parents of typically developing children (29). Such interactional difficulties may contribute to mothers perceiving themselves as less successful in their parenting role, thereby negatively influencing the PA dimension of burnout. In this context, mindful parenting within the parent-child relationship may selectively support the PA dimension of burnout, even in the presence of ADHD-related challenges.

In addition to parenting-related factors, ADHD symptom dimensions demonstrated differential associations with burnout subdimensions. Hyperactivity-impulsivity was independently associated with EE and DP, suggesting that externally disruptive and dysregulating behaviors may place a substantial emotional burden on mothers. Recent studies have provided evidence that externalizing behavioral problems in children may be related to parental burnout (30-32). From an ADHD perspective, hyperactivity and impulsivity are conceptualized as prominent externalizing behaviors (33). In this context, hyperactivity-impulsivity symptoms may be interpreted as child-related behavioral characteristics that are more closely associated with emotionally taxing aspects of maternal burnout, such as EEEE and DP, rather than with mothers' perceptions of PAin their parenting role. In contrast to hyperactivity-impulsivity, which is associated with the emotionally taxing dimensions of burnout, attention-deficit symptoms demonstrate a distinct pattern of association with burnout. Specifically, attention deficit was independently associated with the PA dimension of maternal burnout. Research has also reported that the problems experienced in academic and daily life skills by children with attention deficits are linked to the mother's feeling of being effective and successful in

her role (34). This distinction underscores the importance of considering ADHD symptom dimensions separately when examining the mechanisms of parental burnout. All the children included in the current study were receiving treatment for ADHD at the time of assessment. This ongoing treatment may have contributed to a relative reduction in attention deficit symptoms. This could affect the independent association between attention-deficit symptoms and mothers' perceptions of PA in multivariate models.

This study found no significant relationship between a child's age and the sub-dimensions of maternal burnout. This differs from previous studies that suggest child age has an effect on parental burnout (35,36). That all children in the current sample were undergoing treatment may have masked the effect of children's age on maternal burnout. However, responses to ADHD treatment vary among children (37). In particular, younger children who respond more favorably to treatment may exhibit more manageable behavioral profiles, which could reduce caregiving demands and mitigate maternal burnout. Accordingly, the lack of an age-burnout association in this study may reflect the moderating influence of treatment response rather than the absence of an age-related effect on burnout.

Study Limitations

Although this is the first study to examine the relationship between mindfulness in parenting and burnout in mothers of children with mental disorders, several limitations should be addressed. First, due to its cross-sectional design, it is not possible to make any definite judgments concerning causal relationships. Second, all data relied on mothers' self-reports, and such subjective measurements may entail a risk of bias and measurement error. Third, fathers' or other caregivers' nonparticipation may have resulted in the multifaceted nature of the parenting process not being sufficiently reflected in the study. Additionally, the relatively modest sample size may limit the generalizability of the findings.

Despite these limitations, the present study has several strengths that should be noted. To our knowledge, this is one of the few studies to examine the relationship between mindful parenting and maternal burnout in a clinical sample of mothers of children with ADHD. Furthermore, the dimension-specific and simultaneous assessment of mindful parenting and ADHD symptom profiles provides a more nuanced understanding of burnout mechanisms and may have clinically significant implications for targeted interventions.

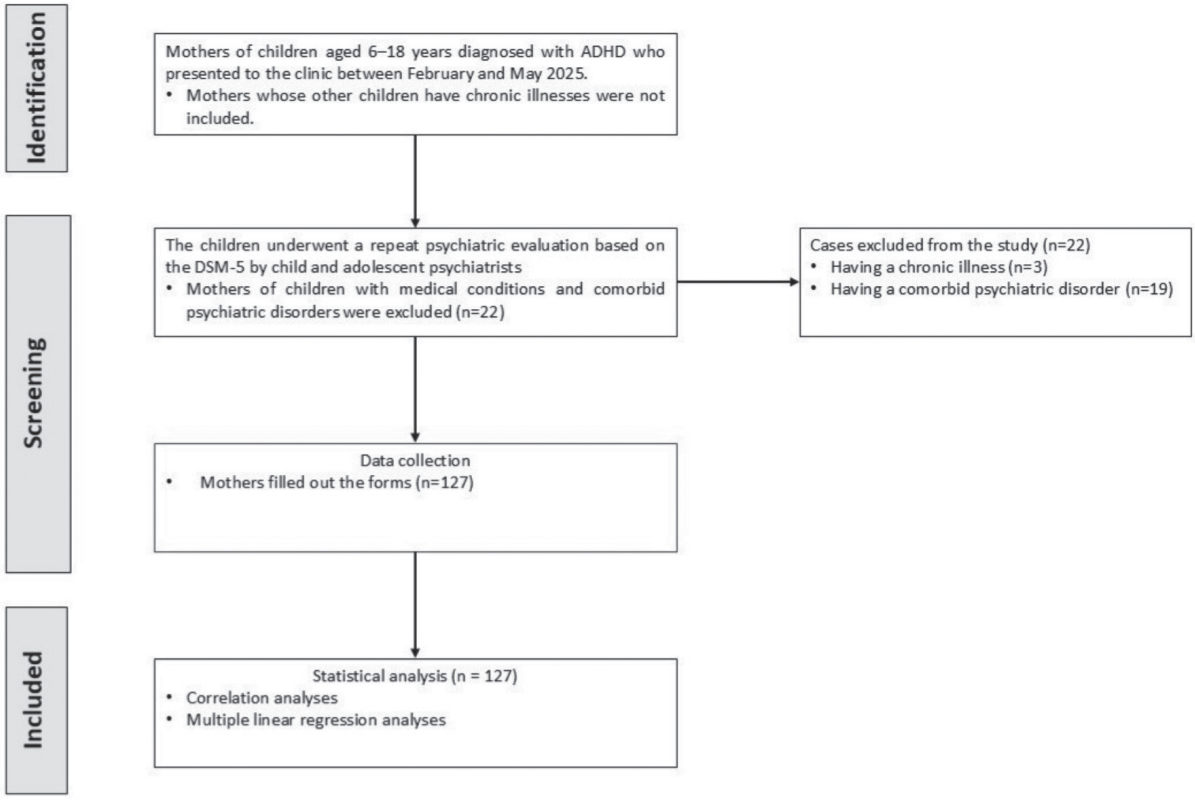

Figure 1. Flowchart of the study
ADHD: Attention deficit hyperactivity disorder

Conclusion

This study revealed significant relationships between mothers' burnout and parenting mindfulness and the severity of their children's ADHD symptoms. The findings show that increased mindfulness in parenting reduces parental burnout. These findings indicate that mindfulness-based parenting intervention programs may be effective in reducing parental burnout. Mindfulness-based approaches and psychoeducational programs aimed at reinforcing parenting experiences may be beneficial in this area. We also believe that further longitudinal and multidimensional studies will yield a more comprehensive analysis of the burnout process by including different parental figures, such as fathers, and external observers, such as teachers or clinicians.

Ethics

Ethics Committee Approval: The study was completed with 127 participants. Approval for the study was granted by the Alanya Alaaddin Keykubat University Faculty of Medicine Clinical Research Ethics Committee (approval no.: 02-08, date: 22.01.2025).

Informed Consent: The purpose of the study was explained to the participants before commencement, and written and verbal consent was obtained from all participants.

Footnotes

Authorship Contributions

Surgical and Medical Practices: O.K., M.S.A., T.K., T.K., P.A.A., Concept: O.K., M.S.A., T.K., Design: O.K., T.K., P.A.A., Data Collection or Processing: O.K., M.S.A., T.K., Analysis or Interpretation: T.K., Literature Search: O.K., P.A.A., Writing: O.K., T.K.

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

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

References

1. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5®). Fifth Edition. Washington DC: American Psychiatric Association; 2013.
2. Salari N, Ghasemi H, Abdoli N, et al. The global prevalence of ADHD in children and adolescents: a systematic review and meta-analysis. Ital J Pediatr. 2023;49:48.
3. Chen L, Liang X, Lu J, et al. Prevalence of attention-deficit hyperactivity disorder (ADHD) and comorbid psychiatric problems in children 12 years and younger: a systematic review and meta-analysis. J Psychiatr Res. 2025;190:547-56.
4. Polanczyk GV, Salum GA, Sugaya LS, Caye A, Rohde LA. Annual research review: a meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. J Child Psychol Psychiatry. 2015;56:345-65.
5. Cunningham CE. A family-centered approach to planning and measuring the outcome of interventions for children with attention-deficit/hyperactivity disorder. J Pediatr Psychol. 2007;32:676-94.
6. Gokcen C, Coskun S, Kutuk MO. Comparison of depression and burnout levels of mothers of children with attention-deficit hyperactivity disorder before and after treatment. J Child Adolesc Psychopharmacol. 2018;28:350-53.
7. Theule J, Wiener J, Tannock R, Jenkins JM. Parenting stress in families of children with ADHD: a meta-analysis Journal of Emotional and Behavioral Disorders. 2013;21:3-17.
8. Urbanowicz AM, Shankland R, Rance J, Bennett P, Gauchet A. Informal mindfulness practices: a new approach to the prevention and treatment of parental burnout. The Family Journal. 2024;32:580-6.
9. Ankori G, Solan M, Plishty S, Brunstein Klomek A, Apter A, Yagil Y. The role of parental qualities in supporting children with ADHD. Children (Basel). 2025;12:845.
10. Duncan LG, Coatsworth JD, Greenberg MT. A model of mindful parenting: implications for parent-child relationships and prevention research. Clin Child Fam Psychol Rev. 2009;12:255-70.
11. Bayot M, Brianda ME, van der Straten N, Mikolajczak M, Shankland R, Roskam I. Treating parental burnout: impact and particularities of a mindfulness- and compassion-based approach. Children (Basel). 2024;11:168.
12. Bögels SM, Lehtonen A, Restifo K. Mindful parenting in mental health care. Mindfulness (N Y). 2010;1:107-20.
13. Zhu H, Shu L, Wang X, Xu Z. Household chaos and preschool migrant children's self-regulation: the mediating role of parent-child conflict and the moderating role of mindful parenting. Front Psychol. 2025;16:1416040.
14. Paucsik M, Urbanowicz A, Leys C, Kotsou I, Baeyens C, Shankland R. Self-compassion and rumination type mediate the relation between mindfulness and parental burnout. Int J Environ Res Public Health. 2021;18:8811.
15. McCaffrey S, Reitman D, Black RA. Mindfulness in parenting questionnaire (MIPQ): development and validation of a measure of mindful parenting. Mindfulness. 2017;8:232-46.
16. Gördesli MA, Arslan R, Çekici F, Sünbül ZA, Malkoç A. The Psyhometric Properties of Mindfulness in Parenting Questionnaire (MIPQ) in Turkish Sample. European Journal of Education Studies. 2018;5:175-88.
17. Maslach C, Jackson SE. The measurement of experienced burnout. Journal of Occupational Behaviour. 1981;2:99-113.
18. Ergin C. Burnout in doctors and nurses and adaptation of Maslach burnout scale. In VII. National Psychology Congress Scientific Studies. 1992;143-54.
19. Burns GL, Preszler J, Becker SP. Psychometric and normative information on the child and adolescent behavior inventory in a nationally representative sample of United States children. J Clin Child Adolesc Psychol. 2022;51:443-52.
20. Çiftçi E, Başay Ö, Özdemir C. Turkish adaptation, validity, and reliability study of the children and adolescent behavior

inventory family questionnaire. Noro Psikiyatr Ars. 2023;60:73-80.

21. George D, Mallery P. IBM SPSS statistics 29 step by step: a simple guide and reference. New York: Routledge; 2024.
22. Chutko LS, Surushkina SY, Yakovenko EA, et al. Kliniko-psikhologicheskie proyavleniya emotsional'nogo vygoraniya u materei detei s psikhonevrologicheskimi zabolevaniyami (na modeli SDVG) [Clinical and psychological manifestations of emotional burnout in mothers of children with neuropsychiatric diseases (on a model of ADHD)]. Zh Nevrol Psikhiatr Im S S Korsakova. 2023;123:97-103. Russian.
23. Findling Y, Itzhaki M, Barnoy S. Parental burnout-a model of risk factors and protective resources among mothers of children with/without special needs. Eur J Investig Health Psychol Educ. 2024;14:2883-900.
24. Scigala KD, Fabris MAF, Zdankiewicz-Ścigała E, Sikora J, Longobardi C. Pandemic era maternal alexithymia and burnout as mediated by self-efficacy and resilience. J Child Fam Stud. 2024;33:2441-54.
25. Muliasari A, Amalia S. Pengaruh dukungan sosial dan parenting self-efficacy terhadap parental burnout pada ibu bekerja. Cognicia. 2024;12:60-6.
26. Ben-Naim S, Gill N, Laslo-Roth R, Einav M. Parental stress and parental self-efficacy as mediators of the association between children's ADHD and marital satisfaction. J Atten Disord. 2019;23:506-16.
27. Duncan LG, Coatsworth JD, Gayles JG, Geier MH, Greenberg MT. Can mindful parenting be observed? Relations between observational ratings of mother-youth interactions and mothers' self-report of mindful parenting. J Fam Psychol. 2015;29:276-82.
28. Baş T, Akgün E. An examination of mothers' self-efficacy and their interaction with their children aged 60-72 months. E-Kafkas Journal of Educational Research 2023;10:629-51.
29. Brown LE, Tallon M, Kendall G, Boyes M, Myers B. Parents' experiences of raising 7- to 11-year-old children with ADHD and perception of a proposed parenting program: a qualitative study. J Atten Disord. 2025;29:312-25.
30. Kerr ML, Fanning KA, Huynh T, Botto I, Kim CN. Parents' self-reported psychological impacts of COVID-19: associations with parental burnout, child behavior, and income. J Pediatr Psychol. 2021;46:1162-71.
31. Yuan Y, Wang W, Song T, Li Y. The mechanisms of parental burnout affecting adolescents' problem behavior. Int J Environ Res Public Health. 2022;19:15139.
32. Chen BB, Qu Y, Yang B, Chen X. Chinese mothers' parental burnout and adolescents' internalizing and externalizing problems: the mediating role of maternal hostility. Dev Psychol. 2022;58:768-77.
33. You Y, McAdams T, Ahmadzadeh YI, Schoeler T, Marzecki F, Zavos HMS. The role of comorbid childhood mental health and neurodevelopmental conditions in the persistence of ADHD symptoms: systematic review and meta-analysis. J Child Psychol Psychiatry. 2026;67:183-201.
34. Clarke AT, Marshall SA, Mautone JA, et al. Parent attendance and homework adherence predict response to a family-school intervention for children with ADHD. J Clin Child Adolesc Psychol. 2015;44:58-67.
35. Pereira MG, Fraga J, Santos M, Ferraz A, Vilaça M. Parental burnout during COVID-19: the moderating role of anxiety and family functioning. Stress Health. 2024;40:e3411.
36. Giraldo CP, Santelices MP, Oyarce D, Chalco EF, Escobar MJ. Children's age matters: parental burnout in Chilean families during the COVID-19 pandemic. Front Psychol. 2022;13:946705.
37. Gul MK, Sener EF, Onal MG, Demirci E. Role of the norepinephrine transporter polymorphisms in atomoxetine treatment: from response to side effects in children with ADHD. J Psychopharmacol. 2022;36:715-22.

Original Article

DOI: 10.4274/haseki.galenos.2026.28247
Med Bull Haseki 2026;64(2):116-121

Psychological Determinants of Vaginal Birth Self-efficacy in Nulliparous Women: The Role of Childbirth Belief Systems

Melike Punduk Yilmaz[1], Iclal Ilknur Ozdemir[2], Berivan Guzelbag[1], Ali Cetin[1]

[1]*University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Obstetrics and Gynecology, Istanbul, Türkiye*

[2]*Istanbul Medipol University Faculty of Health Sciences, Department of Midwifery, Istanbul, Türkiye*

Abstract

Aim: Childbirth self-efficacy is a critical psychological variable that influences birth outcomes and maternal well-being; however, its exact predictors among nulliparous women remain underinvestigated. This research aimed to outline variables that affect self-efficacy for vaginal delivery among nulliparous women, with a specific interest in the effects of birth belief systems.

Methods: This cross-sectional observational study was conducted between February and July 2025 among nulliparous women receiving prenatal care. Validated measures, such as the Birth Beliefs Scale (medical process and natural process subscales) and the Self-efficacy Regarding Vaginal Birth Scale, were used for data collection. Predictors of vaginal birth self-efficacy were determined using hierarchical multiple regression and structural equation modeling.

Results: Two hundred and eighty nulliparous women were included in the study. Self-efficacy was negatively related to medical process beliefs ($b=-0.168$, $p<0.01$) but positively related to natural process beliefs ($b=0.154$, $p<0.01$). The regression model accounted for 18.7% of the variance in self-efficacy [$F(6,273)=10.47$, $p<0.001$]. Other important predictors were planned delivery approach, pregnancy experience rating, education level, and spousal support.

Conclusion: Birth belief systems are the first psychological predictors of vaginal birth self-efficacy in women in their first pregnancy. To help prevent low self-efficacy, healthcare providers should consider adopting screening procedures to identify women whose belief patterns are linked to low self-efficacy.

Keywords: Prenatal education, attitude to health, self-efficacy, parturition, parity

Introduction

Childbirth self-efficacy has been defined as a vital psychological construct that determines birth experiences and obstetric outcomes among nulliparous women (1,2). Self-efficacy is grounded in Bandura's social cognitive theory, which posits that an individual believes in his or her ability to perform certain behaviors when working (3). Empirical evidence shows that increased birth-related self-efficacy is associated with reduced length of labor, better pain management, fewer medical interventions, and greater satisfaction with the birth experience (4,5). Conversely, low self-efficacy is associated with heightened fear of childbirth, increased requests for cesarean section, and higher levels of postpartum psychological distress (6,7).

Among first-time mothers, direct experience is less important for the development of self-efficacy, whereas vicarious learning, verbal persuasion, social support, and messages given by the provider are more relevant

Corresponding Author: Melike Punduk Yilmaz, MD, University of Health Sciences Türkiye, Istanbul Haseki Training and Research Hospital, Clinic of Obstetrics and Gynecology, Istanbul, Türkiye
E-mail: melikepunduk@gmail.com **ORCID:** orcid.org/0000-0001-5942-8026
Received: 23.08.2025 **Accepted:** 26.02.2026 **Publication Date:** 27.03.2026

Cite this article as: Punduk Yilmaz M, Ozdemir II, Guzelbag B, Cetin A. Psychological determinants of vaginal birth self-efficacy in nulliparous women: the role of childbirth belief systems. Med Bull Haseki. 2026;64(2):116-121

(8,9). Another important psychological dimension that potentially influences the development of self-efficacy is the maternal belief system regarding childbirth (10). An orientation toward medical processes defines birth as a risk-taking process requiring technological intervention and professional care, whereas an orientation toward natural processes emphasizes maternal ability and the normalcy of birth (11,12). Such belief orientations are found to shape the birth preferences, labor coping mechanisms, and final birth outcomes (13,14).

Although previous studies emphasize the importance of psychological factors for childbirth outcomes, few systematic studies have examined the relationship between childbirth belief systems and self-efficacy, especially among nulliparous women who have no prior experience in childbirth (15,16). These relationships are important to understand in order to develop specific prenatal interventions likely to increase maternal confidence. However, evidence evaluating the simultaneous influence of childbirth belief systems and socio-demographic factors on vaginal birth self-efficacy among nulliparous women remains limited, particularly in middle-income countries. We hypothesized that vaginal birth self-efficacy would be significantly predicted by birth belief orientations, with medical-process beliefs showing negative correlations and natural-process beliefs showing positive correlations. In this context, we aimed to explore the determinants of vaginal birth self-efficacy in nulliparous women, with a specific focus on childbirth beliefs compared with demographic and clinical factors.

Materials and Methods

Complicance with Ethical Standarts

The research was approved by the Istanbul Medipol University Non-Interventional Clinical Research Ethics Committee (approval no.: 1360, date: 26.12.2024). The data were collected at Haseki Training and Research Hospital through an online survey. The electronic written informed consent of all participants was obtained in accordance with the principles of the Declaration of Helsinki.

Design of the Study and Study Population

The proposed cross-sectional observational study was conducted to examine factors affecting self-efficacy for vaginal birth among nulliparous women recruited from prenatal care centers between February and July 2025. The study design was based on the STROBE guidelines for observational studies (17). The inclusion criteria were nulliparous women; gestational age 16-38 weeks; singleton pregnancy; and no history of miscarriage or termination of pregnancy. Exclusion criteria were multiple gestation, known fetal anomalies, and severe maternal medical conditions requiring specialized obstetric treatment. G*Power software was used to compute the sample size, with the following parameters: multiple regression analysis (effect size f^2=0.15, alpha =0.05, power =0.95), which returned a minimum required sample size of 172 participants. Figure 1 is the flow diagram of the study.

Data Collection Instruments

The demographic variables recorded were maternal age, educational level, employment status, income level, and family structure. Clinical measurements included gestational age, body mass index, high-risk pregnancy, spousal support, and lifestyle variables, including physical activity and nutrition.

The Birth Beliefs Scale is a scale that measures beliefs about birth processes. It was created by Preis and Benyamini (10) and validated in Turkish by Paker and Ertem (18). The scale comprises two subscales: the medical process subscale (6 items), which includes beliefs that the birth process must be medically performed and controlled by experts; and the natural process subscale (5 items), which includes beliefs that women have the ability and physiological normality of the birth process. The internal consistency is acceptable in both subscales (Cronbach's alpha =0.80).

Self-efficacy with vaginal birth scale, created by Chu et al. (19) and translated into the Turkish language by Karadeniz and Kavlak (20), is used to assess maternal self-confidence in their ability to handle labor pain, adhere to medical instructions, and maintain control over their emotions during the process of giving birth, using 9 items. Its scale shows good internal consistency (Cronbach's alpha =0.92).

Statistical Analysis

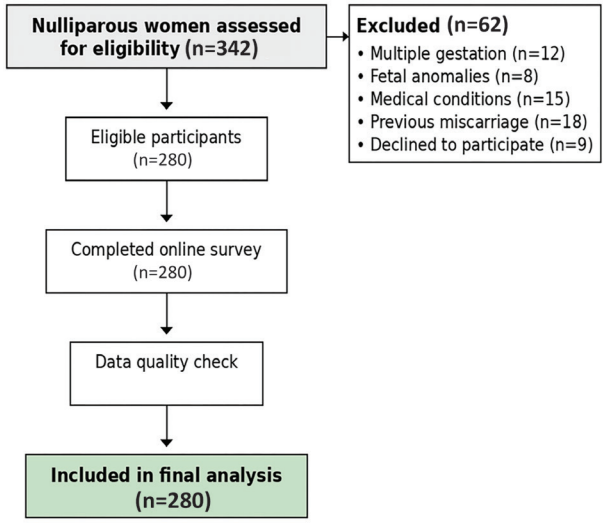

Figure 1. Study flow diagram

R software version 4.3.2 was used to carry out statistical analyses (21). Descriptive statistics included means (with standard deviations) for continuous variables and frequencies (with percentages) for categorical variables. Normality was assessed using Shapiro-Wilk tests and Q-Q plots. Pearson correlation was used to analyze bivariate relationships between two continuous variables; Spearman's correlation was used to analyze bivariate relationships between two ordinal variables; point-biserial correlation was used to analyze bivariate relationships between two binary variables; and the eta coefficient was used to analyze bivariate relationships between two nominal variables. The enter method was used in a multiple linear regression analysis, with self-efficacy as the dependent variable. Multicollinearity was evaluated using variance inflation factors, with a threshold of 5. Structural equation model (SEM) was used to investigate theoretical associations using the lavaan package (22). The comparative fit index (CFI 0.95), Tucker-Lewis index (TLI 0.95), root-mean-square error of approximation (RMSEA <0.06), and standardized root-mean-square residual (SRMR <0.08) were used to evaluate model fit. Statistical significance was set at p=0.05.

Results

Table 1 shows descriptive statistics. The nulliparous women recruited in the study numbered 280, and the mean age of the study population was 27.40±3.90. Most of them completed tertiary education (62.5%), were in skilled or professional jobs (72.1%), and reported middle-to-high income levels (63.6%). The majority (86.8%) had low-risk pregnancies, and most respondents (85.4%) rated their pregnancy experience as positive. The mean scores for medical process beliefs, natural process beliefs, and vaginal birth self-efficacy were 3.00±0.76, 3.81±1.04, and 59.14±20.71, respectively. These findings reveal moderately high support for medical-process beliefs, moderately high support for natural-process beliefs, and moderately high self-efficacy among the study population.

The bivariate analysis showed significant associations with self-efficacy (Table 2). The beliefs about medical process were negatively correlated (r=-0.199, p<0.01), but the beliefs about natural process were positively correlated (r=0.189, p<0.01). The level of education (r=0.142, p=0.05), pregnancy experience rating (r=0.156, p=0.05), spousal support (rpb=0.134, p=0.05), and planned delivery approach (r=0.198, p=0.01) were also significantly correlated with self-efficacy.

Table 1. Descriptive characteristics of study participants (n=280)

Variable	Mean±SD / n (%)
Demographic variables	
Age (years)	27.40±3.90
Higher education	175 (62.5%)
Skilled/professional employment	202 (72.1%)
Middle-high income	178 (63.6%)
Nuclear family	251 (89.6%)
Pregnancy-related variables	
Gestational age (weeks)	25.18±5.87
Low-risk pregnancy	243 (86.8%)
Good/very good pregnancy experience	239 (85.4%)
BMI (kg/m²)	26.91±4.49
Spousal support present	230 (82.1%)
Childbirth education received	136 (48.6%)
Natural vaginal delivery planned	201 (71.8%)
Scale scores	
Medical process beliefs (1-5)	3.00±0.76
Natural process beliefs (1-5)	3.81±1.04
Vaginal birth self-efficacy (0-90)	59.14±20.71
SD: Standard deviation, BMI: Body mass index	

Table 2. Associations between study variables and vaginal birth self-efficacy (n=280)

Variable	Effect size	95% CI	p-value
Medical process beliefs	r=-0.199	(-0.314, -0.078)	<0.01
Natural process beliefs	r=0.189	(0.073, 0.301)	<0.01
Education level	ρ=0.142	(0.025, 0.255)	<0.05
Pregnancy experience	ρ=0.156	(0.040, 0.268)	<0.05
Spousal support	rpb=0.134	(0.017, 0.247)	<0.05
Planned delivery approach	ρ=0.198	(0.121, 0.275)	<0.01
Age	r=-0.084	(-0.199, 0.034)	NS
Gestational age	r=0.079	(-0.038, 0.194)	NS
BMI	r=0.087	(-0.030, 0.202)	NS

Pearson correlation was used for continuous variables, Spearman's correlation for ordinal variables, point-biserial correlation for binary variables, and eta coefficient for nominal variables
r: Pearson correlation, ρ: Spearman correlation, rpb: Point-biserial correlation, η: Eta coefficient, CI: Confidence interval, NS: Not significant, BMI: Body mass index

The multiple regression model was statistically significant and accounted for 18.7% of the variance in self-efficacy scores [F(6,273)=10.47, p<0.001]. The strongest negative predictor was medical process beliefs (b=0.168, p<0.01), and the strongest positive predictor was natural process beliefs (b=0.154, p<0.01). These findings are presented in Table 3.

The SEM showed a good fit (CFI=0.962, TLI=0.951, RMSEA=0.048, SRMR=0.065). The results of the path analysis indicated that self-efficacy was indirectly affected by education level via natural process beliefs, which mediated the relationship between demographic factors and maternal confidence.

Discussion

This research investigated the psychological and socio-demographic factors influencing vaginal birth self-efficacy in nulliparous women. The results indicated that childbirth belief systems were significant predictors of maternal self-efficacy. In particular, beliefs about medical processes were linked to lower self-efficacy, while beliefs about natural processes were linked to higher maternal confidence.

The current literature defines the belief systems about birth as key psychological factors that influence self-efficacy for vaginal birth among nulliparous women. The results of this study provide empirical support for theoretical frameworks that focus on psychological elements in predicting birth experience (23,24) and supplement the current literature on modifiable influences on maternal confidence.

The negative relationship between medical process beliefs and self-efficacy (b=-0.168) is clinically significant. It raises the possibility that an excessive emphasis on medical procedures can unintentionally undermine maternal confidence in the ability to give birth naturally. Women who strongly favor medical interventions may become dependent on external medical care and lose confidence in their physiological ability to give birth normally (25). This finding is consistent with previous studies that indicate the possibility of diminished maternal agency and increased anxiety during labor associated with excessive medicalization of birth procedures (26,27). Healthcare providers are advised to keep in mind that, although relevant medical information is necessary, the tendency to present childbirth as a medical event that requires intervention might adversely affect maternal self-efficacy.

On the other hand, the natural process beliefs and self-efficacy were positively related (b=0.154), a finding that contributes to interventions based on maternal competence and physiological birth. This result is consistent with studies showing that women who believe that birth is a normal process have higher self-efficacy and better birth outcomes (28,29). The SEM analysis found that natural process beliefs act as mediators that help translate educational benefits into increased maternal confidence. Positive messages emphasizing women's inherent birthing abilities that are included in prenatal education programs could also be effective in improving self-efficacy (30,31).

The results of our study agree with recent research showing that psychological variables such as birth beliefs and delivery preferences have a greater impact on maternal confidence than conventional demographic or clinical variables (32,33). The 18.7 per cent explained variance in beliefs about birth shows that birth beliefs are significant factors in the development of self-efficacy, although other factors need to be explored. Further studies should examine how fear of childbirth, perceived quality of social support networks, and healthcare providers' communication styles can influence the development of maternal self-efficacy (34,35).

Study Limitations

This research has several limitations that are worth considering. The cross-sectional design precludes drawing causal conclusions because self-efficacy and birth beliefs can influence each other over time. Directional relationships would be better supported by longitudinal studies that observe changes in beliefs and self-efficacy during pregnancy. The use of a single-center recruitment

Table 3. Multiple regression analysis predicting vaginal birth self-efficacy (n=280)

Predictor variable	β	SE	95% CI	p-value	VIF
Medical process beliefs	-0.168	0.042	(-0.251, -0.085)	<0.01	1.89
Natural process beliefs	0.154	0.038	(0.079, 0.229)	<0.01	1.76
Planned delivery approach	0.134	0.041	(0.053, 0.215)	<0.01	2.34
Pregnancy experience	0.112	0.032	(0.049, 0.175)	<0.01	1.34
Education level	0.089	0.028	(0.034, 0.144)	<0.05	1.23
Spousal support	0.078	0.036	(0.007, 0.149)	<0.05	1.12
Model summary	R^2=0.187		F(6,273) =10.47	<0.001	

Multiple linear regression analysis using enter method was applied
β: Standardized beta coefficient, SE: Standard error, CI: Confidence interval, VIF: Variance inflation factor

approach can limit external validity in other care or cultural environments. Use of self-report measures may lead to response bias, such as social desirability effects. Self-efficacy formation may be influenced by unmeasured variables such as fear of childbirth, perceived quality of support, pre-existing exposure to birth stories, and relationships with healthcare providers. In addition, the relatively uniform socio-economic traits of the study population may restrict the applicability of the results to more heterogeneous groups, such as those with varying income levels, educational backgrounds, and cultural beliefs about childbirth. Lastly, the clinical implications of self-efficacy on birth outcomes could not be directly assessed, as behavioral outcomes such as the actual delivery method were not investigated.

Nevertheless, this study provides new evidence regarding psychological factors associated with vaginal birth self-efficacy among Turkish nulliparous women. The study's overall statistical approach, using multiple regression and SEM, strengthens the plausibility of the observed relationships. The use of validated Turkish instruments ensures that the measures are culturally appropriate and valid. The sample size was adequate and exceeded the minimum a priori sample size calculated to ensure sufficient statistical power for the intended analyses.

Conclusion

Birth belief systems are key psychological antecedents of self-efficacy for vaginal birth among nulliparous women. Medical process beliefs are negatively correlated with maternal confidence, whereas natural process beliefs are positively correlated with maternal confidence. Healthcare providers should consider screening women to identify those with belief patterns associated with low self-efficacy, enabling the provision of individualized prenatal care that maximizes maternal confidence and perhaps enhances birth outcomes. Prenatal education programs that balance required medical content with positive messages about women's birthing abilities may effectively increase the self-efficacy of nulliparous women. Therefore, targeting childbirth beliefs through prenatal education should be a strategy to improve maternal confidence and potentially promote vaginal birth.

Declaration Regarding the Use of AI and AI-Assisted Technologies: During the preparation of this work, the authors used artificial intelligence tools to assist in verifying statistical analyses and formatting the manuscript. After carefully reviewing and editing the content as necessary, the authors take full responsibility for the publication's content.

Ethics

Ethics Committee Approval: The research was approved by the Istanbul Medipol University Non-Interventional Clinical Research Ethics Committee (approval no.: 1360, date: 26.12.2024).

Informed Consent: All participants provided electronic written informed consent prior to participation.

Footnotes

Authorship Contributions

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

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

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

References

1. Lowe NK. Self-efficacy for labor and childbirth fears in nulliparous pregnant women. J Psychosom Obstet Gynaecol. 2000;21:219-24.
2. Carlsson IM, Ziegert K, Nissen E. The relationship between childbirth self-efficacy and aspects of well-being, birth interventions and birth outcomes. Midwifery. 2015;31:1000-7.
3. Bandura A. Self-efficacy: toward a unifying theory of behavioral change. Psychol Rev. 1977;84:191-215.
4. Frankham LJ, Thorsteinsson EB, Bartik W. Childbirth self-efficacy and birth related PTSD symptoms: an online childbirth education randomised controlled trial for mothers. BMC Pregnancy Childbirth. 2024;24:668.
5. Schwartz L, Toohill J, Creedy DK, Baird K, Gamble J, Fenwick J. Factors associated with childbirth self-efficacy in Australian childbearing women. BMC Pregnancy Childbirth. 2015;15:29.
6. Beebe KR, Lee KA, Carrieri-Kohlman V, Humphreys J. The effects of childbirth self-efficacy and anxiety during pregnancy on prehospitalization labor. J Obstet Gynecol Neonatal Nurs. 2007;36:410-8.
7. Salomonsson B, Gullberg MT, Alehagen S, Wijma K. Self-efficacy beliefs and fear of childbirth in nulliparous women. J Psychosom Obstet Gynaecol. 2013;34:116-21.
8. Alizadeh-Dibazari Z, Abdolalipour S, Mirghafourvand M. The effect of prenatal education on fear of childbirth, pain intensity during labour and childbirth experience: a scoping review using systematic approach and meta-analysis. BMC Pregnancy Childbirth. 2023;23:541.
9. Ip WY, Tang CS, Goggins WB. An educational intervention to improve women's ability to cope with childbirth. J Clin Nurs. 2009;18:2125-35.
10. Preis H, Benyamini Y. The birth beliefs scale - a new measure to assess basic beliefs about birth. J Psychosom Obstet Gynaecol. 2017;38:73-80.
11. Chadwick RJ, Foster D. Negotiating risky bodies: childbirth and constructions of risk. Health, Risk & Society. 2014;16:68-83.

12. Van der Gucht N, Lewis K. Women's experiences of coping with pain during childbirth: a critical review of qualitative research. Midwifery. 2015;31:349-58.
13. Aktaş Reyhan F, Dağli E. The effect of positive birth story videos on women's fear of birth, childbirth self- efficacy and birth preference: a randomized controlled study. Women & Health. 2025;65:559-70.
14. Sucu S, Sucu ST, Soysal Ç. Investigating the influence of antenatal education on birth beliefs and delivery methods: a prospective cohort study from Turkey. BMC Pregnancy Childbirth. 2025;25:491.
15. Ji KM, Li ZZ, Min H, Sun LP, Zhao Y. The trajectory and influencing factors of fear of childbirth in third trimester primiparas: a prospective longitudinal study. J Adv Nurs. 2025;81:4712-22.
16. Preis H, Pardo J, Peled Y, Benyamini Y. Changes in the basic birth beliefs following the first birth experience: self-fulfilling prophecies? PLoS ONE. 2018;13:e0208090.
17. von Elm E, Altman D, Egger M et al. The Strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. The Lancet. 2007;370:1453-7.
18. Paker S, Ertem G. Validity and reliability study of the Turkish form of the birth beliefs scale. Anatol J Gen Med Res. 2022;32:1-8.
19. Chu KH, Chen AC, Tai CJ, Chen SF, Chien LY. Development and validation of the self-efficacy regarding vaginal birth scale. J Obstet Gynecol Neonatal Nurs. 2017;46:e13-21.
20. Karadeniz H, Kavlak O. Validity and reliability of the turkish version of the self-efficacy regarding vaginal birth scale: a methodological study. Value Health Sci. 2024;14:416-24.
21. R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2023.
22. Rosseel Y. lavaan: an R package for structural equation modeling. Journal of Statistical Software. 2012;48:1-36.
23. Fenwick J, Toohill J, Creedy DK, Smith J, Gamble J. Sources, responses and moderators of childbirth fear in Australian women: a qualitative investigation. Midwifery. 2015;31:239-46.
24. Stoll K, Edmonds JK, Hall WA. Fear of childbirth and preference for cesarean delivery among young American women before childbirth: a survey study. Birth. 2015;42:270-6.
25. Shaw D, Guise JM, Shah N, et al. Drivers of maternity care in high-income countries: can health systems support woman-centred care? Lancet. 2016;388:2282-95.
26. Davis-Floyd R. The technocratic, humanistic, and holistic paradigms of childbirth. Int J Gynaecol Obstet. 2001;75:S5-23.
27. Johansson M, Rubertsson C, Rådestad I, Hildingsson I. Childbirth - an emotionally demanding experience for fathers. Sex Reprod Healthc. 2012;3:11-20.
28. Larkin P, Begley CM, Devane D. Women's experiences of labour and birth: an evolutionary concept analysis. Midwifery. 2009;25:e49-59.
29. Leap N, Sandall J, Buckland S, Huber U. Journey to confidence: women's experiences of pain in labour and relational continuity of care. J Midwifery Womens Health. 2010;55:234-42.
30. Fritzson M, Nordin A, Børøsund E, Johansson M, Varsi C, Ängeby K. A mobile application for early labour support -feasibility pilot study. Women and Birth. 2023;36:495-503.
31. Gagnon AJ, Sandall J. Individual or group antenatal education for childbirth or parenthood, or both. Cochrane Database Syst Rev. 2007;2007:CD002869.
32. Nilsson C, Lundgren I, Karlström A, Hildingsson I. Self reported fear of childbirth and its association with women's birth experience and mode of delivery: a longitudinal population-based study. Women Birth. 2012;25:114-21.
33. Rouhe H, Salmela-Aro K, Gissler M, Halmesmäki E, Saisto T. Mental health problems common in women with fear of childbirth. BJOG. 2011;118:1104-11.
34. Hildingsson I, Haines H, Karlström A, Nystedt A. Presence and process of fear of birth during pregnancy-Findings from a longitudinal cohort study. Women Birth. 2017;30:e242-7.
35. Toohill J, Fenwick J, Gamble J, Creedy DK. Prevalence of childbirth fear in an Australian sample of pregnant women. BMC Pregnancy Childbirth. 2014;14:275.

Original Article

DOI: 10.4274/haseki.galenos.2026.96967
Med Bull Haseki 2026;64(2):122-128

Association Between Prognostic Nutritional Index and Major Amputation in Patients with Diabetic Foot Ulcers

Basak Can, Esra Deniz Kahvecioglu, Ibrahim Ethem Tas, Seyda Nur Kurtuncuoglu

University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital, Clinic of Internal Medicine, Istanbul, Türkiye

Abstract

Aim: Given the clinical importance of early risk stratification in patients with diabetic foot ulcers and the limited availability of simple and reliable biomarkers, there is a need for easily accessible parameters that can support clinical decision-making. Therefore, this study aimed to evaluate the association between the Prognostic Nutritional Index (PNI) and major amputation in patients hospitalized for diabetic foot ulcers.

Methods: This retrospective observational study included patients with diabetic foot ulcers evaluated by the Diabetic Foot Board between January 2020 and August 2024, who were managed conservatively or underwent minor or major amputation. Demographic characteristics, comorbid diseases, and laboratory parameters—including C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), HbA1c, neutrophil-to-lymphocyte ratio (NLR), systemic immune-inflammation index (SII), and PNI—were compared among groups. Multivariable logistic regression analysis was performed to identify independent factors associated with major amputation, and receiver operating characteristic curve analysis was used to assess the discriminative performance of PNI.

Results: The study cohort consisted of 756 cases: 216 without amputation, 338 with minor amputation, and 202 with major amputation. Significant differences were observed between the amputation and non-amputation groups with respect to age, coronary artery disease, chronic heart failure, CRP, ESR, NLR, SII, and PNI. Age, ESR, CRP, and PNI were identified as independently associated with major amputation.

Conclusion: The PNI reflects both immunological and nutritional status and is independently associated with major amputation in patients with diabetic foot ulcers.

Keywords: Prognostic nutritional index, amputation, immunonutrition, inflammation

Introduction

The global prevalence of diabetes mellitus in adults has surpassed 800 million and continues to increase, making it a major public health concern (1). Diabetes is associated with several microvascular and macrovascular complications, including nephropathy, neuropathy, retinopathy, and peripheral arterial disease (2). Among these, the diabetic foot ulcer is one of the most devastating complications, resulting from the combined effects of neuropathy, ischemia, impaired immunity, and chronic inflammation, and leading to delayed wound healing and increased susceptibility to severe infection. Despite advances in multidisciplinary management, 15-20% of diabetic foot ulcers still require amputation, and major amputation is associated with markedly increased mortality and reduced quality of life (3).

The Prognostic Nutritional Index (PNI) is a peripheral blood–based immunonutritional marker calculated using the serum albumin level and the lymphocyte count, reflecting both nutritional and immune status. Malnutrition, systemic inflammation, and immune dysfunction play key roles in the progression of diabetic foot ulcers; therefore,

Corresponding Author: Basak Can, MD, University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital, Clinic of Internal Medicine, Istanbul, Türkiye
E-mail: basakozkurt@gmail.com **ORCID:** orcid.org/0000-0003-1910-9922
Received: 07.11.2025 **Accepted:** 04.03.2026 **Publication Date:** 27.03.2026

Cite this article as: Can B, Kahvecioglu ED, Tas IE, Kurtuncuoglu SN. Association between prognostic nutritional index and major amputation in patients with diabetic foot ulcers. Med Bull Haseki. 2026;64(2):122-128

PNI may represent the biological vulnerability to severe outcomes such as major amputation (4,5). However, evidence regarding the association between PNI and amputation in diabetic foot disease is limited, and its specific relationship with major amputation remains unclear. We hypothesized that a low PNI is independently associated with an increased risk of major amputation in patients with diabetic foot disease.

The aim of this study was to evaluate the association between PNI and major amputation in patients with diabetic foot disease. By identifying a simple and easily accessible biomarker for early risk stratification, this study may contribute to improved clinical decision-making and the timely implementation of preventive and therapeutic strategies in high-risk patients.

Materials and Methods

Compliance with Ethical Standards

The study protocol received approval from the University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital Scientific Research Ethics Committee No. 1 (approval no.: KAEK/28.08.2024.187, date: 17.09.2024). All patients who fulfilled the inclusion criteria and were managed by the board between January 2020 and August 2024 were enrolled.

Study Design

This study was designed as a retrospective observational study based on the medical records of patients with diabetic foot ulcers evaluated by the Diabetic Foot Board between January 2020 and August 2024. The Diabetic Foot Board decided on management for patients, including minor amputation, major amputation, and non-amputation treatments such as medical therapy and vascular interventions. Based on treatment outcomes, participants were grouped into three categories: minor amputation, major amputation, or non-amputation management.

Treatment decisions were made by a multidisciplinary Diabetic Foot Board consisting of vascular surgeons, infectious disease specialists, endocrinologists, orthopedic surgeons, and plastic surgeons. Decisions were based on a structured clinical assessment, including ulcer depth and extent, presence and severity of infection, vascular status (clinical examination, ankle-brachial index, and imaging findings when available), radiological evidence of osteomyelitis, and feasibility of revascularization. Major amputation was performed when the limb was considered non-salvageable due to extensive necrosis, uncontrolled infection, or irreversible ischemia.

The study flowchart is presented in Figure 1. In our study, amputations below the ankle level were classified as minor, while those above the ankle were considered major. Amputations at the ankle joint were classified as major because they significantly affect walking ability.

Patients aged >80 years were excluded to minimize confounding by advanced frailty, multimorbidity, and age-related alterations in albumin synthesis and in lymphocyte counts that could bias PNI interpretation. Individuals who had chronic liver disease affecting albumin levels or whose data were incomplete were not included in the study. Patients using medications that affect hematological parameters were excluded. Patients who declined the amputation recommended by the Diabetic Foot Board were excluded from the final analysis, as outcome classification was based on the actual treatment performed. The patients' demographic data and medical histories, including comorbidities such as rheumatologic disease, peripheral neuropathy, venous insufficiency, coronary artery disease, chronic heart failure, chronic kidney disease, arterial thrombosis, venous thrombosis, peripheral artery disease, and osteomyelitis, were recorded. Peripheral artery disease was defined based on clinical history, an ankle-brachial index of <0.9, or vascular imaging findings (6).

Laboratory parameters were obtained at hospital admission, prior to the Diabetic Foot Board evaluation, as part of a routine clinical assessment. Due to the urgent nature of some admissions, fasting status was not standardized; however, all measurements were performed in the hospital's central laboratory using standard procedures. Laboratory parameters obtained prior to the Board evaluation—including C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), HbA1c, neutrophil-to-lymphocyte ratio, systemic immune-inflammation index (SII), and PNI—were recorded and compared among the groups. Prognostic Nutritional Index was calculated as [10×serum albumin (g/dL)] + [0.005×lymphocyte count (/mm^3)], and SII was calculated as (platelet×neutrophil)/lymphocyte count.

In our retrospective study design, the outcome (major amputation, minor amputation, or non-amputation treatment) was determined at the time of the Diabetic Foot Board decision. Therefore, there was no additional follow-up period beyond the board decision.

Statistical Analysis

Statistical analyses were performed using IBM SPSS Statistics version 21.0 (IBM Corp., Armonk, NY, USA). Continuous variables were expressed as mean ± standard deviation for normally distributed data and as median (minimum-maximum) for non-normally distributed data, as appropriate. The distribution of continuous variables was assessed using the Shapiro-Wilk test, supported by visual inspection of histograms and Q-Q plots.

Categorical variables were compared using the chi-square test or Fisher's exact test, as appropriate. For comparisons of continuous variables among three independent groups, the Kruskal-Wallis test was used for non-normally distributed data. When a significant difference was detected, post hoc pairwise comparisons were performed using the Mann-Whitney U test with Bonferroni correction.

Multivariable logistic regression analysis was conducted to identify factors independently associated with major amputation. Model calibration was evaluated using the Hosmer-Lemeshow goodness-of-fit test, with a p-value>0.05 indicating an adequate model fit. Multicollinearity among independent variables was assessed using the variance inflation factor (VIF), and no significant collinearity was observed (all VIF values were <2). The adequacy of the multivariable logistic regression model was evaluated based on the events-per-variable (EPV) ratio. With 202 major amputation events and 8 candidate predictors included in the model, the EPV was approximately 25, exceeding the commonly recommended minimum threshold of 10 and the more conservative threshold of 20, thereby indicating a low risk of model over-fitting.

To control for type I error in multiple pairwise comparisons among the three groups (non-amputation vs. minor amputation, non-amputation vs. major amputation, and minor amputation vs. major amputation), Bonferroni correction was applied, and the adjusted significance level was set at α=0.0167. A two-tailed p-value<0.05 was considered statistically significant.

Results

A total of 756 patients managed by the Diabetic Foot Board were enrolled in the study. Of these, 216 (28.5%) were in the non-amputation group, 338 (44.8%) were in the minor amputation group, and 202 (26.7%) were in the major amputation group.

The patients' demographic characteristics and medical histories, including comorbidities, are summarized in Table 1. The mean age was 59.7±10.9 years in the non-amputation group, 61.5±10.9 years in the minor amputation group, and 63.7±10.0 years in the major amputation group. The age of patients in the major

Table 1. Baseline demographic and clinical characteristics of the study groups

	Non-amputation n=216 (28%)	Minor amputation n=338 (45%)	Major amputation n=202 (%27)	p-value
Age	59.7±10.9	61.5±10.9	63.7±10	**0.002**[ξ]
Gender Male/Female	164/52	253/85	142/60	0.376[*]
Cigarette Yes/No	65/150	117/221	74/127	0.342[*]
Rheumatologic disease Yes/No	5/211	5/333	5/197	0.669[#]
Peripheral neuropathy Yes/No	66/150	127/211	66/136	0.203[*]
Venous insufficiency Yes/No	14/202	21/317	10/192	0.787[*]
Coronary artery disease Yes/No	111/105	164/173	125/77	**0.01**[*]
Chronic heart failure Yes/No	62/154	95/242	79/123	**0.019**[*]
Chronic kidney disease Yes/No	60/154	101/237	73/129	0.168[*]
Arterial thrombosis Yes/No	62/154	88/250	72/130	0.058[*]
Venous thrombosis Yes/No	7/209	9/329	8/194	0.722[*]
Peripheral artery disease Yes/No	187/29	300/38	179/23	0.724[*]
Osteomyelitis Yes/No	138/78	237/100	135/67	0.28[*]

[*]Chi-square test, [#]Fisher's exact test, [ξ]Kruskal-Wallis test.
Age; Non-amputees vs. Minor amputees p=0.082, Non-amputees vs. Major amputees p<0.0001, Minor amputees vs. Major amputees p=0.036
Coronary artery disease; Non-amputees vs. Minor amputees p=0.532, Non-amputees vs. Major amputees p=0.031, Minor amputees vs. Major Amputees p=0.003
Chronic heart failure; Non-amputees vs. Minor amputees p=0.859, Non-amputees vs. Major amputees p=0.03, Minor amputees vs. Major amputees p=0.009

amputation group was significantly higher than that of the other two groups (non-amputation vs. Minor amputation, p=0.082; Non-amputation vs. Major amputation, p<0.0001; Minor amputation vs. Major amputation, p=0.036).

The numbers of patients with coronary artery disease in the non-amputation, minor amputation, and major amputation groups were 111 of 216, 164 of 338, and 125 of 202, respectively. The prevalence of coronary artery disease was significantly greater in the major amputation group than in the other two groups (non-amputation vs. Minor amputation, p=0.532; Non-amputation vs. Major amputation, p=0.031; Minor amputation vs. Major amputation, p=0.003).

The numbers of patients with chronic heart failure were 62 of 216, 95 of 338, and 79 of 202 in the non-amputation, minor amputation, and major amputation groups, respectively. The prevalence of chronic heart failure was significantly higher in the major amputation group than in the other two groups (Non-amputation vs. Minor amputation, p=0.859; Non-amputation vs. Major amputation, p=0.03; Minor amputation vs. Major amputation, p=0.009).

The PNI and hematological values of the groups are presented in Table 2. A comparison of hematological parameters and indices revealed that both the minor and major amputation groups exhibited significantly higher levels of CRP, ESR, neutrophil-to-lymphocyte ratio, and SII compared with the non-amputation group, whereas their PNI was significantly lower (p<0.001).

The multivariable logistic regression model is shown in Table 3. The Hosmer–Lemeshow test indicated good model calibration (χ^2=2.01, p=0.98), with no significant difference between observed and predicted outcomes. Logistic regression analysis identified advanced age, elevated CRP and ESR levels, and lower PNI as significant risk factors for major amputation.

The results of the receiver operating characteristic (ROC) analysis, including sensitivity and specificity values illustrating the predictive performance of the PNI for major amputation, are summarized in Table 4. A PNI cut-off value of 34 predicted major amputation with a

Table 2. Inflammatory markers and prognostic nutritional index across study groups

	Non-amputees n=216 (28%)	Minor amputees n=338 (45%)	Major amputees n=202 (27%)	p-value	Non-amp. vs. Minor	Non-amp. vs. Major	Minor vs. Major
CRP (mg/L) median (min-max)	29 (0.2-321)	46 (0.8-407)	117 (0.7-468)	<0.001λ	0.003	<0.001	<0.001
ESR (mm/h) median (min-max)	58 (2-132)	62 (4-140)	82 (1-140)	<0.001λ	0.052	<0.001	<0.001
HbA1c (%) median (min-max)	8.2 (4-15)	8.6 (5-15)	8 (4-13)	0.09λ	-	-	-
Neutrophil/lymphocyte ratio median (min-max)	3.46 (0.1-37)	4 (0.3-107)	6.2 (1.1-51)	<0.00λ	0.211	<0.001	<0.001
Systemic immune inflammatory index median (min-max)	1186 (40-20692)	1466 (185-30391)	2651 (283-19157)	<0.001λ	0.008	<0.001	<0.001
Prognostic nutritional index median (min-max)	36 (3-49)	37 (19-49)	30 (16-74)	<0.001λ	0.698	<0.001	<0.001

λKruskal-Wallis test
CRP: C-reactive protein, ESR: Erythrocyte sedimentation rate

Table 3. Multivariable logistic regression analysis of factors associated with major amputation

	Odds ratio (OR)	95% CI (clinically meaningful)	p-value
Age (per 1 year)	1.029	1.010-1.048	**0.002**
Coronary artery disease	0.918	0.581-1.474	0.722
Chronic heart failure	0.720	0.449-1.156	0.173
Peripheral artery disease	1.078	0.75-1.230	0.800
CRP (per 10 mg/L)	1.046	1.020-1.072	<0.001
ESR (per 10 mm/h)	1.111	1.041-1.185	**0.001**
Neutrophil/lymphocyte ratio	0.971	0.921-1.024	0.277
Systemic immune inflammatory index	1.000	0.998-1.043	0.089
Prognostic nutritional index (per 5 points)	0.650	0.56-0.76	**<0.001**

CRP: C-reactive protein, ESR: Erythrocyte sedimentation rate, CI: Confidence interval

Table 4. The findings of the ROC analysis with sensitivity, specificity values demonstrating the success of the prognostic nutritional index in major amputation prediction

AUC (95% CI)	Cut-off	p-value	Sensitivity %	Specificity %
0.750 (0.711-0.789)	34	<0.001	70.3%	70.2%

ROC: Receiver operating characteristic, AUC: Area under the curve, CI: Confidence interval

sensitivity of 70.3% and a specificity of 70.2% (p<0.001). The corresponding ROC curve for the PNI is depicted in Figure 2.

Discussion

We found a significant association between lower PNI levels and the risk of major amputation in patients with diabetic foot ulcers. Diabetic foot ulcers are recognized as one of the late complications associated with diabetes. Over time, patients with diabetes may develop peripheral motor and sensory neuropathies. This neuropathy leads to a reduced perception of pain and temperature, causing minor wounds or pressure points on the foot to go unnoticed. With the addition of other contributing factors, diabetic foot ulcers develop (7). While some individuals with diabetic foot ulcers are managed with medical treatments, approximately 20% undergo lower extremity amputation, either minor or major (8). Although losing a limb is challenging for any patient, predictably, major amputations have more severe consequences. In this context, identifying factors influencing major amputation is crucial.

The relationship between PNI and diabetes complications has been previously investigated. Recent studies have indicated that the PNI serves as a risk factor for both the progression of diabetic nephropathy and increased mortality (9,10). A different study demonstrated a relationship between PNI and increased mortality risk in elderly patients with chronic kidney disease (11).

Prognostic Nutritional Index is calculated using a formula that includes serum albumin levels and lymphocyte count (12). Albumin is a marker of a patient's nutritional status. Low albumin levels indicate malnutrition and increased disease severity (13). Malnutrition in individuals with diabetes mellitus is associated with an increased risk of macrovascular complications. Additionally, protein plays a crucial role in the tissue healing process, and its deficiency can lead to delayed wound healing (14). Moreover, protein deficiency due to malnutrition can impair immune response mechanisms (15). Therefore, the association between hypoalbuminemia—a marker of malnutrition—and amputation risk may be explained by its impact on the immune system and its contribution to vascular complications. Although albumin is a component of PNI and low albumin plausibly contributes to poor wound healing, albumin was not analyzed separately in this study; therefore, causal statements about albumin alone cannot be made from our data. Another parameter used to calculate PNI is the lymphocyte count. A low lymphocyte count may indicate dysfunction or suppression of the immune system (16). In the diabetic foot, bacterial infections often lead to increases in leukocyte and neutrophil counts, while lymphocytes and albumin tend to decrease as negative acute-phase reactants. These examples indicate that decreases in albumin and lymphocytes and, consequently, in PNI increase the risk of amputation.

A study evaluating the relationship between PNI and total amputation (both minor and major) found that the predictive value of PNI for forecasting amputation was 39 (17). In our study, the PNI had a predictive cut-off of 34 for major amputation. The discrepancy likely reflects differences in outcome definition (any amputation vs. major only) and patient characteristics. In clinical practice, a PNI<34 could identify patients who may benefit from early nutrition consultation and who should be prioritized for revascularization or antibiotic therapy. This finding suggests that PNI may be used not only as a prognostic marker in clinical decision-making but also as a risk stratification tool to guide the intensity of treatment.

Alongside PNI, other indicators of systemic inflammation and general physiological susceptibility—such as advanced age, heightened ESR, and elevated CRP—were independently correlated with major amputation in our study. These findings indicate that both immunonutritional status and inflammatory burden influence limb outcomes in diabetic foot disease. As individuals age, the prevalence of atherosclerosis and peripheral artery disease rises, resulting in compromised blood circulation and protracted tissue healing. The lack of a notable difference between groups in peripheral artery disease may indicate the substantial baseline prevalence of vascular disease within this specific diabetic foot cohort, coupled with the extensive diagnostic criteria employed in standard clinical practice. Also, age-related immunosenescence leads to impaired immune function, thereby increasing susceptibility to infection. Because of this, issues like deep, infected ulcers and osteomyelitis can happen, which may mean that a limb needs to be cut off. It is well known that ESR and CRP, two inflammatory markers, are often much higher in these situations. Recent research indicates that elevated inflammatory markers, such as ESR and CRP, correlate with a heightened risk of major amputation (18). A recent study examining predictive factors for amputation in diabetic

feet revealed that elevated CRP levels and advanced age are significant risk factors for amputation (19). Our results align with existing literature, demonstrating that elevated ESR and CRP levels, along with advanced age, correlate with a heightened risk of major amputation. While inflammatory markers like CRP and ESR may be linked to parts of PNI, the multivariable regression analysis indicated that PNI was still linked to major amputation even after controlling for age and inflammatory factors. This indicates that PNI encompasses supplementary prognostic data beyond singular inflammatory markers.

Interestingly, HbA1c levels did not significantly differ among groups. This finding suggests that acute inflammatory burden and vascular status may play a more immediate role in amputation decisions than long-term glycemic control. While poor glycemic control contributes to ulcer development, acute infection severity and ischemia may be more decisive in determining major amputation.

Study Limitations

Several limitations should be acknowledged. First, the study's results may not apply to other situations since it was done at only one center. Second, the study's retrospective design carries an inherent risk of selection bias. Third, the lack of a validated system for classifying ulcer severity, like the Wagner or University of Texas classification systems, is a big problem. Ulcer depth, infection severity, and ischemia are well-established determinants of amputation risk, and the lack of standardized severity stratification may have led to residual confounding. Furthermore, comprehensive data concerning revascularization status and procedural outcomes were excluded from the analysis, potentially exacerbating unmeasured confounding. Consequently, PNI ought to be regarded as a supplementary biomarker rather than a replacement for established clinical staging systems. Lastly, the study did not look at mortality outcomes, which made it harder to figure out what would happen to patients in the long term.

Despite these limitations, this study has several strengths, including a well-defined patient cohort, a thorough immunonutritional assessment using PNI, and the evaluation of clinically significant outcomes such as major amputation. These factors offer significant insights into the prognostic value of PNI in patients with diabetic foot disease and may inform future risk stratification and management approaches.

Conclusion

The Prognostic Nutritional Index reflects both immune function and nutritional status and is associated with major amputation in patients with diabetic foot ulcers. Evaluation of immunonutritional status may provide additional insight into limb prognosis in patients with diabetic foot ulcers. The PNI, as an accessible and cost-effective marker, may serve as a complementary tool for risk stratification and multidisciplinary decision-making.

Ethics

Ethics Committee Approval: The study protocol received approval from the University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital Scientific Research Ethics Committee No. 1 (approval no.: KAEK/28.08.2024.187, date: 17.09.2024).

Informed Consent: Informed consent was obtained from the patients.

Footnotes

Authorship Contributions

Surgical and Medical Practices: B.C., E.D.K., Concept: B.C., E.D.K., Design: B.C., E.D.K., Data Collection or Processing: B.C., E.D.K., I.E.T., S.N.K., Analysis or Interpretation: B.C., E.D.K., I.E.T., S.N.K., Literature Search: B.C., E.D.K., I.E.T., S.N.K., Writing: B.C.

Conflict of Interest: The authors declare that they have no conflict of interest related to this study.

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

References

1. Ong KL, Stafford LK, McLaughlin SA, et al. Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021. Lancet. 2023;402:203-34.
2. Dunn TJ, Tan X, Harton J, et al. Macrovascular and microvascular complications in US medicare enrollees with type 2 diabetes with and without atherosclerotic cardiovascular disease. Diabetes, Obes Metab. 2025;27:4137-47.
3. Lu Q, Wang J, Wei X, Wang G, Xu Y. Risk factors for major amputation in diabetic foot ulcer patients. Diabetes Metab Syndr Obes. 2021;14:2019-27.
4. Noronha JC, Mechanick JI, Barazzoni R, et al. Malnutrition, sarcopenia and nutrition therapy for patients with diabetes - a general framework and focus on hospital care. Clin Nutr ESPEN. 2025;70:8-17.
5. Rashid T, Zia S, Mughal S, Baloch AA, Abdul Rauf MU, Hasan SM. Prevalence of malnutrition and associated factors among the elderly with type 2 diabetes using MNA form. J Nutr Metab. 2025;2025:2107146.
6. Ghirardini F, Martini R. Current opinion on diagnosis of peripheral artery disease in diabetic patients. Medicina (Kaunas). 2024;60:1179.
7. Dawi J, Tumanyan K, Tomas K, et al. Diabetic foot ulcers: pathophysiology, immune dysregulation, and emerging therapeutic strategies. Biomedicines. 2025;13:1076.

8. Liu Z, Wei D, Wang J, Gao L. Predicting major amputation risk in diabetic foot ulcers using comparative machine learning models for enhanced clinical decision-making. Sci Rep. 2025;15:28103.
9. Zhang J, Xiao X, Wu Y, et al. Prognostic nutritional index as a predictor of diabetic nephropathy progression. Nutrients. 2022;14:3634.
10. Zhang J, Chen Y, Zou L, Gong R. Prognostic nutritional index as a risk factor for diabetic kidney disease and mortality in patients with type 2 diabetes mellitus. Acta Diabetol. 2023;60:235-45.
11. Barutcu Atas D, Tugcu M, Asicioglu E, et al. Prognostic nutritional index is a predictor of mortality in elderly patients with chronic kidney disease. Int Urol Nephrol. 2022;54:1155-62.
12. Onodera T, Goseki N, Kosaki G. Prognostic nutritional index in gastrointestinal surgery of malnourished cancer patients. Nihon Geka Gakkai Zasshi. 1984;85:1001-5. [Japanese]
13. Margarson MP, Soni N. Serum albumin: touchstone or totem? Anaesthesia. 1998;53:789-803.
14. Xu S, Wang Y, Hu Z, Ma L, Zhang F, Liu P. Effects of neutrophil-to-lymphocyte ratio, serum calcium, and serum albumin on prognosis in patients with diabetic foot. Int Wound J. 2023;20:1638-46.
15. Chandra RK. Nutrition and the immune system: an introduction. Am J Clin Nutr. 1997;66:460S-3S.
16. Pendersen BK, Toft AD. Effects of exercise on lymphocytes and cytokines. Br J Sports Med. 2000;34:246-51.
17. Coşkun B, Ayhan M, Ulusoy S. Relationship between prognostic nutritional index and amputation in patients with diabetic foot ulcer. Diagnostics. 2024;14:738.
18. Sharma N, Tandup C, Rastogi A, Sahu S, Behera A, Savlania A. Predictors of major lower extremity amputation in type 2 diabetic patients with diabetic foot ulcers: a cross-sectional analytical study. Cureus. 2025;17:1-7.
19. Farine F, Rapisarda AM, Roani C, et al. Predictive factors of amputation in diabetic foot. Biomedicines. 2024;12:1-11.

Original Article

DOI: 10.4274/haseki.galenos.2026.96158
Med Bull Haseki 2026;64(2):129-137

Association Between Thyroid Nodules and Thyroid Cancer Risk in Graves' Disease: A Surgical Cohort from an Iodine-deficient Region

Tugba Matlim Ozel[1], Mine Yilmaz[1], Sezer Akbulut[1], Aykut Celik[1], Gorkem Yildiz[1], Nilsen Yildirim Erdogan[2], Serkan Sari[1]

[1]*University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital, Clinic of General Surgery, Istanbul, Türkiye*
[2]*University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital, Clinic of Pathology, Istanbul, Türkiye*

Abstract

Aim: The clinical relevance of thyroid nodules in Graves' disease (GD) remains controversial, particularly in iodine-deficient regions where both nodularity and thyroid cancer are more prevalent. This study aimed to compare clinical, biochemical, ultrasonographic, cytological, and pathological characteristics between nodular and non-nodular GD (non-NGD) to clarify the oncologic significance of nodularity.

Methods: This retrospective observational cohort study included 160 patients who underwent total thyroidectomy for GD between June 2020 and July 2025 at an endocrine surgery center. Patients were classified according to preoperative ultrasonography (US) as nodular or non-NGD. Demographic features, thyroid autoimmunity markers, ophthalmopathy, fine-needle aspiration biopsy (FNAB) results, and final pathology were recorded. Multivariate logistic regression analysis was performed to identify independent predictors of nodularity.

Results: Sixty-three patients (39.4%) had nodular GD (NGD), and 97 patients (60.6%) had non-NGD. Nodular patients were older and had higher body mass index (BMI) (both $p<0.05$), whereas thyroid-stimulating immunoglobulin and anti-thyroid peroxidase levels and ophthalmopathy were significantly higher in non-nodular patients. Fine-needle aspiration biopsy was more frequently performed in the nodular group (57.1% vs. 20.6%). Overall, papillary thyroid carcinoma (PTC) was diagnosed in 31.8% of the cohort, with a markedly higher prevalence in NGD (52.4% vs. 18.5%, $p<0.001$). Tumors in nodular patients were larger and more likely to exhibit lymphatic invasion. In multivariate analysis, age and BMI remained independent predictors of nodularity.

Conclusion: In this surgically treated GD cohort, nodularity identifies a structural phenotype with substantially increased PTC risk, while non-NGD reflects an autoimmune-dominant phenotype. High malignancy rates across Bethesda categories and even in non-biopsied patients indicate the need for vigilant US surveillance and a low threshold for FNAB of suspicious or dominant nodules.

Keywords: Graves disease, nodular graves disease, thyroid nodule, papillary thyroid carcinoma, biopsy, fine-needle, hyperthyroidism

Introduction

Graves' disease (GD) is the most common cause of autoimmune hyperthyroidism and remains a major global health issue with marked geographical variability in presentation, comorbidities, and treatment practices (1,2). Although GD is encountered worldwide, its clinical profile may be significantly influenced by regional iodine status, access to specialized endocrine care, and institutional preferences for therapy (3). In iodine-deficient regions such as Türkiye, nodular thyroid disease is more prevalent, underscoring the importance of careful evaluation for synchronous nodules in GD, since both nodularity and

Corresponding Author: Lec, Tugba Matlim Ozel, MD, University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital, Clinic of General Surgery, Istanbul, Türkiye
E-mail: tugbamatlim79@gmail.com **ORCID:** orcid.org/0000-0001-6305-2957
Received: 02.01.2026 **Accepted:** 04.03.2026 **Publication Date:** 27.03.2026

Cite this article as: Matlim Ozel T, Yilmaz M, Akbulut S, et al. Association between thyroid nodules and thyroid cancer risk in Graves' disease: a surgical cohort from an iodine-deficient region. Med Bull Haseki. 2026;64(2):129-137

iodine-deficiency-related thyroid hyperplasia may influence cancer risk and clinical decision-making (4,5).

While the thyroid gland is usually diffusely enlarged in GD, nodules can also be present and detected through imaging studies (6,7). With the increased use of high-resolution ultrasonography (US), thyroid nodules are now detected in 25-50% of GD patients worldwide (2,8). In iodine-deficient regions, this prevalence may be even higher (4,5). As nodularity is a well-recognized determinant of malignancy risk in euthyroid individuals, its presence in GD has attracted considerable research interest. Over the last decade, contemporary meta-analyses have consistently demonstrated that GD patients with nodules have a 3- to 5-fold increased risk of thyroid cancer (TC) compared with nodule-free GD patients (9-11). Importantly, several regional studies from endemic goiter areas have similarly reported elevated cancer rates among nodular GD (NGD) patients, with prevalence figures exceeding those seen in iodine-sufficient countries (5).

Globally, differentiated thyroid carcinoma (DTC) incidence has risen substantially, driven partly by advances in imaging and more frequent detection of small carcinomas (12). This trend is reflected in GD cohorts, where the majority of detected malignancies are papillary thyroid carcinomas (PTC) (10,11,13). However, some iodine-deficient or endemic goiter populations have demonstrated higher rates of multifocality, tall-cell variant tumors, or other aggressive pathological characteristics among GD-associated cancers (5,14). These findings may reflect the combined effects of chronic iodine deficiency, thyrotropin receptor antibody (TRAb)-mediated proliferative stimulation, autoimmune inflammation, and oxidative DNA damage—all of which have been implicated in GD-related oncogenesis (3,12,15).

Despite the accumulating evidence, knowledge gaps persist regarding the precise clinical implications of nodularity in GD. Preoperative US characteristics, fine-needle aspiration biopsy (FNAB) accuracy, nodule size thresholds, Thyroid-stimulating immunoglobulin (TSI)/autoimmunity profiles, and the pathological behavior of cancers in nodular versus non-nodular GD (non-NGD) remain incompletely understood, and studies vary widely in methodology and patient selection (2,11,13). In the literature, various rates have been reported regarding the prevalence of thyroid nodules in GD, and the malignant potential of these nodules remains controversial (2,9).

We hypothesized that the presence of thyroid nodules in GD identifies a distinct clinical phenotype associated with a higher risk of malignancy and with specific tumor characteristics. Accordingly, this study aimed to evaluate the prevalence and oncologic significance of thyroid nodules in GD by comparing nodular and non-nodular patients in a large surgical cohort from an iodine-deficient region and to provide insights that may support more individualized and region-specific management strategies.

Materials and Methods

Compliance with Ethical Standards

This study was approved by the University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital Institutional Ethics Committee (approval no: 463, date: 17.12.2025) and was conducted in accordance with the Declaration of Helsinki. Due to the retrospective nature of the study and the use of anonymized data, the Ethics Committee waived the requirement for written informed consent.

Study Design and Setting

This retrospective observational cohort study was conducted at the Division of Endocrine Surgery, Department of General Surgery, University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital, Istanbul, Türkiye. The study included patients with GD who underwent total thyroidectomy at a tertiary, high-volume academic center between June 2020 and July 2025. Clinical and perioperative data were prospectively recorded during routine clinical care and retrospectively analyzed for the purposes of this study.

Patient Selection and Grouping

Eligible patients met the diagnostic criteria for GD based on clinical examination, biochemical hyperthyroidism, positive TSI, and typical findings on scintigraphy (diffuse uptake) and/or on US. Exclusion criteria included less than TT, age <18 years, prior thyroid surgery, previous headandneck malignancy, final pathological diagnosis of TC other than PTC (including 2 follicular TC, 1 medullary TC, and 1 oncocytic TC), or missing data. The flow diagram of patient selection is shown in Figure 1. Patients were categorized according to preoperative US:

- **NGD group:** Rresence of solid or mixed nodules,
- **non-NGD group*:*** Absence of synchronous thyroid nodules.

Preoperative Evaluation and Surgery

Preoperative assessment included thyroid function tests, thyroid autoantibodies, highresolution US, and selective FNAB of nodules with suspicious features. All ultrasonographic examinations and FNABs were performed by experienced radiologists specialized in thyroid imaging. Thyroid nodule cytopathology was classified according to the 2023 Bethesda System for Reporting Thyroid Cytopathology (16). Nodule characteristics and FNABhistopathology concordance were recorded. All procedures were performed by a dedicated endocrine surgery team. Euthyroidism was achieved preoperatively using antithyroid drugs (ATDs), beta-

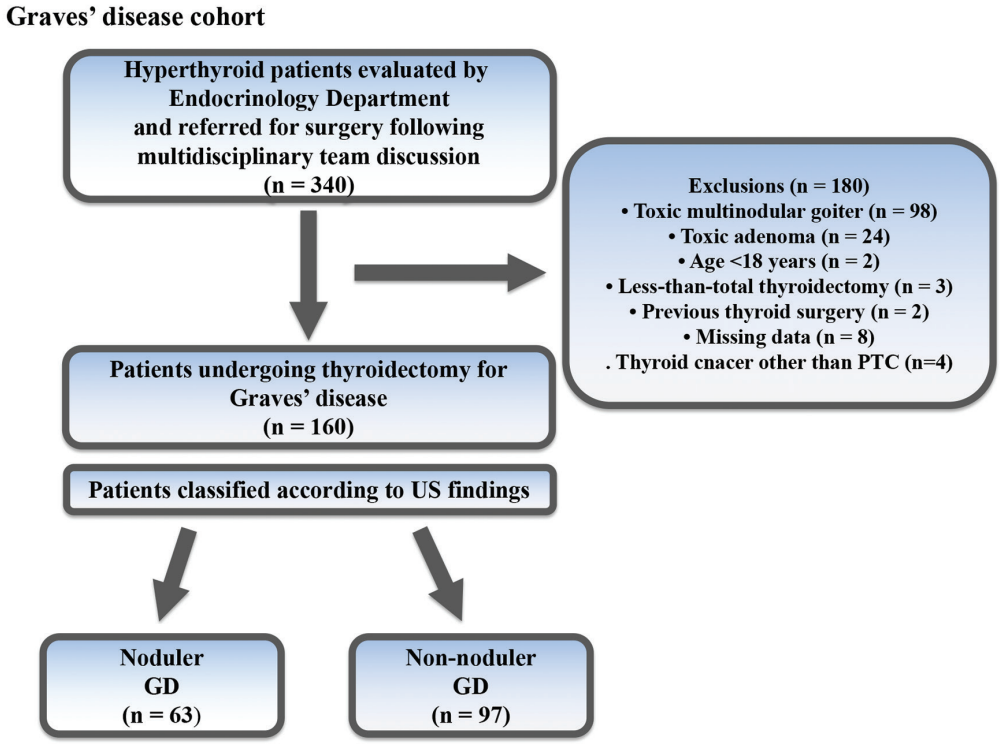

Figure 1. Flow diagram illustrating patient selection, exclusion criteria, and final study group allocation
GD: Graves' disease, US: Ultrasonography

blockers (β-blockers), and Lugol's solution when indicated. Surgical indications included relapse or intolerance to medical therapy; failure or relapse after radioactive iodine treatment (RAI); compressive symptoms; suspicion of malignancy; indeterminate cytology; planning pregnancy; contraindications to RAI; and patient preference. Central lymph node dissection was performed selectively when clinical or radiologic suspicion existed. Thyroid-stimulating hormone (TSH), FT3, FT4, and anti-thyroid peroxidase (anti-TPO) were measured in serum samples using a sandwich electrochemiluminescence immunoassay on the COBAS 8000 e801 analyzer (Roche Diagnostics, Mannheim, Germany). Thyroid-stimulating immunoglobulin levels were measured using a chemiluminescence immunoassay. The normal values for these assays are TSH 0.27-4.2 uIU/mL, FT4 0.93-1.7 ng/dL, FT3 2.6-4.4 pg/mL, anti-TPO 0-34 IU/mL, and TSI <0.1 (0.1-0.55) U/L.

Histopathological Evaluation

Surgical specimens were evaluated by the same pathology team in accordance with the World Health Organization 2022 diagnostic criteria (17). Assessed parameters included tumor size, variant, multifocality, bilaterality, lymphatic or vascular invasion, capsular invasion, extrathyroidal extension (ETE), and lymph node metastasis. Papillary microcarcinoma was defined as ≤10 mm.

Variables and Outcomes

Collected variables included demographics [age, sex, and body mass index (BMI)]; preoperative use of Graves' medications (methimazole, propylthiouracil, β-blockers steroids, Lugol solution, or potassium iodide); diagnosis of ophthalmopathy; thyroid function and autoimmunity markers; US findings; nodule characteristics; cytology results; and histopathology. Primary outcomes:

- Nodule prevalence in GD,
- Comparison of clinical, biochemical, and imaging variables between nodular and nonNGD,
- Malignancy rates and tumor behavior in relation to nodularity.

Statistical Analysis

The primary statistical objective was to identify independent predictors of nodular disease using multivariate logistic regression analysis. Continuous variables were expressed as mean ± standard deviation or median (interquartile range), and categorical variables as n (%). Normality of continuous variables was assessed using the Shapiro-Wilk test. Between-group comparisons were performed using Student's t-test, Mann-Whitney U test, or chi-square test, as appropriate. Variables with $p<0.10$ in univariate analysis or considered clinically relevant based on prior literature were included in the

multivariate logistic regression model. Odds ratios with 95% confidence intervals were reported. Statistical analyses were performed using SPSS, version 25.0 (IBM Inc., Armonk, NY, USA), and p<0.05 was considered statistically significant. Because the study objective was to compare nodular and non-NGD phenotypes within a surgical cohort, nodularity was defined as the dependent variable in the multivariate regression model. A separate multivariate model with malignancy as the dependent variable was not constructed.

Results

Comparison Between Nodular and non-NGD

A total of 160 patients who underwent TT for GD were included. Of these, 63 patients (39.4%) had at least one thyroid nodule on preoperative US, while 97 patients (60.6%) had a diffusely enlarged, non-nodular gland. Patients in the nodular group were significantly older than those without nodules (43.2±11.4 vs. 38.6±11.4 years, p=0.01). Body mass index was also higher in the nodular group (28.85±5.68 vs. 25.82±4.56 kg/m², p<0.001). In contrast, markers of autoimmune activity were more prominent in the non-nodular group: median TSI (8.27 vs. 3.41, p=0.002) and anti-TPO levels (107 vs. 31 IU/mL, p=0.03) were significantly higher. Ophthalmopathy was also more frequent in the non-nodular group (69% vs. 52.4%, p=0.04). Preoperative steroid therapy (43.3% vs. 28.6%, p=0.087) and plasmapheresis (7.2% vs. 0%, p=0.01) tended to be more frequent among non-nodular patients, while ATDs and RAI use were comparable. There was no significant difference in sex distribution between patients with and without nodules (female proportion: 31.7% vs. 30.9%; p=0.99). The duration of preoperative medical therapy was similar between groups. Mean duration was 18.19±9.95 months in the nodular group and 22.62±8.45 months in the non-nodular group (p=0.24) (Table 1).

Preoperative Cytology and FNAB Utilization

Fine-needle aspiration biopsy was performed significantly more often in patients with nodules (57.1% vs. 20.6%, p<0.001). Among FNAB-sampled nodular patients (n=36), the Bethesda II, III-IV, and V-VI categories accounted for 52.8%, 19.4%, and 27.8%, respectively. In non-NGD patients undergoing preoperative FNAB (n=20), Bethesda II cytology accounted for 80% of cases, while Bethesda III-IV and Bethesda V-VI each comprised 10% (Table 2).

FNAB-histopathology Concordance

Malignancy rates across Bethesda categories varied between groups: Bethesda II nodules showed malignancy in 42% of nodular patients vs. 12.5% of non-nodular patients; for Bethesda III-IV, 28.5% vs. 50%; and for Bethesda V-VI, 90% vs. 100%. Importantly, malignancy was also present in patients who had not undergone FNAB: 16.8% (13/77) in non-NGD and 51.9% (14/27) in NGD (Table 2).

Malignancy Rates and Tumor Pathology

Overall, 51 of 160 patients (31.8%) were diagnosed with PTC. The malignancy rate was significantly higher in the nodular group (52.4%) than in the non-nodular group (18.5%) (p<0.001). Tumor size was significantly larger

Table 1. Comparison of demographic, clinical findings between nodular and non-nodular GD

Variables	Non-nodular GD (n=97)	Nodular GD (n=63)	p-value
Age (year), mean ± SD	38.58±11.4	43.2±11.4	**0.01***
Gender, n (%) Female Male	30 (30.9) 67 (69.1)	20 (31.7) 43 (68.3)	0.99‡
Preoperative BMI, mean ± SD**	25.82±4.56	28.85±5.68	**<0.001***
Preoperative medical therapy period (month), mean ± SD	22.62±8.45	18.19±9.95	0.24*
Preoperative TSH, median (Q1-Q3)	0.05(0.02-0.82)	0.34(0.05-1.18)	0.07†
Preoperative anti-TPO, median (Q1-Q3)	107(25.4-284)	31(9-295)	**0.03†**
Preoperative TSI, median (Q1-Q3)	8.27(3.11-18.4)	3.41(1.21-7.34)	**0.002†**
Preoperative medical therapy use (n, %) ATD Steroid RAI Plasmapheresis	97 (100) 42 (43.3) 9 (9.3) 7 (7.2)	62 (98.5) 18 (28.6) 6 (9.5) 0 (0)	0.39*** 0.087‡ 1.000‡ **0.01****
Ophthalmopathy presence (n, %)	67 (69)	33 (52.4)	**0.04‡**

*Student's t-test, †Mann-Whitney U test, ‡Chi-square test, ***Fisher's exact test, bold values indicate statistically significant results (p<0.05)
**BMI: Body mass index, GD: Graves' disease, TSH: Thyroid-stimulating hormone, anti-TPO: Anti-thyroid peroxidase, TSI: Thyroid-stimulating immunoglobulin, ATD: antithyroid drug, RAI: radioactive iodine treatment, SD: Standard deviation

in nodular patients (median 7 mm vs. 3 mm, p=0.007). Lymphatic invasion was more common in nodular patients (54.5% vs. 22.2%, p=0.03), while vascular invasion, ETE, multifocality, bilaterality, and lymph-node metastasis were similar between the groups (Table 2).

Predictors of Nodule Presence

Univariate comparisons demonstrated that NGD was more common among older, higher-BMI patients and was accompanied by increased preoperative FNAB use and a greater postoperative detection rate of PTC. Ophthalmopathy was inversely associated with the presence of nodules (OR: 0.49, p=0.049). In the multivariate model, age (OR 1.04, p=0.04) and BMI (OR 1.11, p=0.01) remained independent predictors of nodular disease, whereas anti-TPO, TSI, steroid therapy, plasmapheresis, and ophthalmopathy were not independently associated with nodular disease (Table 3).

Discussion

The markedly higher prevalence of malignancy observed in NGD in our cohort supports the concept that nodularity represents a structural disease phenotype associated with thyroid carcinogenesis rather than a coincidental morphological finding. Contemporary surgical series have similarly demonstrated that TC in GD occurs predominantly in glands exhibiting nodular remodeling rather than diffuse autoimmune enlargement alone (1,5,8,11,18).

Multiple large series and meta-analyses consistently indicate that the increased TC risk observed in GD is largely confined to patients with nodular glandular architecture rather than to those with diffuse autoimmune enlargement. Both Papanastasiou et al. (9) and Huang and Chen (10) demonstrated that the excess carcinoma risk in GD is predominantly attributable to coexisting nodularity, while Palella et al. (11) further confirmed NGD as a high-risk subgroup in their umbrella analysis (9-11). This collective evidence supports the interpretation that nodularity represents a structural disease phenotype

Table 2. Comparison of histopathological findings between nodular and non-nodular GD

Variables	Non-nodular GD (n=97)	Nodular GD (n=63)	p-value
Preoperative FNAB n (%) **No** **Yes**	 77 (79.4) 20 (20.6)	 27 (42.9) 36 (57.1)	**<0.001‡**
Preoperative Bethesda category n (%) **Bethesda II** **Bethesda III and IV** **Bethesda V and VI**	 16 (80) 2 (10) 2 (10)	 19 (52.8) 7 (19.4) 10 (27.8)	0.126‡
Final malignancy rate according to FNAB n (%) **No FNAB** **Bethesda II** **Bethesda III and IV** **Bethesda V and VI**	 13 /77 (16.8) 2/16 (12.5) 1/2 (50) 2/2 (100)	 14/27 (51.9) 8/19 (42) 2/7 (28.5) 9/10 (90)	
Papillary thyroid cancer rate (n, %)	18/97 (18.5)	33/63 (52.4)	**<0.001‡**
Papillary microcarcinoma rate (n,%)	16/18 (88.9)	22/33 (66.7)	0.16‡
Tumor diameter, median (Q1-Q3)	3 (2-5.5)	7 (4.5-12)	**0.007†**
Tumor subtype, n (%) **Classic** **Foliculer** **Tallcell** **Hobnail** **Others**	 14 (77.8) 2 (11.1) 2 (11.1) 0 (0) 0 (0)	 18 (54.5) 6 (18.2) 5 (15.2) 1 (3) 3 (9.1)	0.49‡
Lymphatic invasion, n (%)	4 (22.2)	18 (54.5)	**0.03‡**
Multifocal disease, n (%)	4 (22.2)	13 (39.4)	0.21‡
Bilaterality, n (%)	2 (11.2)	10 (30.3)	0.12‡
Extrathyroidal extension, n (%)	1 (5.6)	1 (3)	0.99‡
Vascular invasion, n (%)	0 (0)	4 (12.1)	0.28‡
Lymph node metastasis, n (%)	1 (5.6)	7 (21.2)	0.23‡

†Mann-Whitney U test, ‡Chi-square test, bold values indicate statistically significant results (p<0.05)
FNAB: Fine-needle aspiration biopsy, GD: Graves' disease

Table 3. Univariate and multivariate analysis of factors associated with preoperative ultrasonographic nodularity in Graves' disease (nodular vs non-nodular GD)

	Univariate analysis†			Multivariate analysis		
Variables	**OR**	**95% CI**	**p-value**	**OR**	**95% CI**	**p-value**
Age	1.04	1.007-1.07	**0.02**	1.04	1.002-1.08	**0.04**
BMI	1.13	1.05-1.21	**0.001**	1.11	1.03-1.21	**0.01**
Preoperative TSI	0.97	0.94-1.001	0.06	0.99	0.96-1.03	0.66
Preoperative anti-TPO	0.999	0.997-1.001	0.28	1	0.99-1.001	0.81
Ophthalmopathy presence	0.49	0.25–0.95	0.049	0.78	0.37–1.63	0.51
Steroid use	0.52	0.27–1.03	0.087	0.60	0.27-1.33	0.21
Plasmapheresis	0.09	0.01–0.74	**0.01**	0.14	0.02-1.10	0.06
Preoperative FNAB situation	5.13	2.55-10.34	**<0.001**	4.97	2.17-11.35	**<0.001**
Papillary TC†	4.83	2.37-9.84	**<0.001**	3.56	1.56-8.15	**0.003**

Bold values indicate statistically significant results (p<0.05)
†Papillary thyroid carcinoma status was included as an associated pathological variable rather than a causal predictor of nodularity
GD: Graves' disease, TSI: Thyroid-stimulating immunoglobulin, anti-TPO: Anti-thyroid peroxidase, TC: Thyroid cancer, OR: Odds ration, CI: Confidence Interval, BMI: Body mass index, FNAB: Fine-needle aspiration biopsy,

associated with malignant transformation in GD rather than merely a coincidental morphological finding. In our cohort, the markedly higher cancer prevalence observed in NGD aligns with this structural risk model, suggesting that glandular remodeling and nodular transformation are key determinants of carcinogenic potential in GD.

Our malignancy rate in NGD exceeds pooled global estimates but parallels observations from endemic goiter regions, particularly Turkish surgical series in which Keskin et al. (8), Kefeli et al. (5), and Dayanan et al. (19) consistently emphasized nodularity as the principal determinant of malignancy in GD. These region-specific data suggest that in iodine-deficient settings, TC in GD is more similar to malignancy arising in nodular thyroid disease than to diffuse autoimmune goiter. Consequently, referral and selection patterns may enrich the surgical population with higher-risk disease and limit direct comparison with international series derived from iodine-sufficient cohorts or from non-surgical GD cohorts. In contrast, cohorts from iodine-sufficient populations have reported lower cancer frequencies and questioned whether GD itself—independent of nodular transformation—confers meaningful oncologic risk (20). This geographic divergence likely reflects differences in iodine status, glandular remodeling, and surgical referral patterns rather than the intrinsic biological heterogeneity of GD. In this context, our findings further support the concept that nodularity defines a high-risk structural phenotype within GD, particularly in endemic regions where both nodular transformation and carcinogenic potential are amplified. These observations reinforce the importance of meticulous ultrasonographic evaluation, maintaining a low threshold for FNAB, and considering earlier definitive surgical management in selected NGD patients.

Our findings support a two-phenotype model in GD, distinguishing an autoimmune-dominant, non-nodular form from a structurally remodeled, nodular phenotype associated with increased malignant potential. This conceptual framework is consistent with regional clinical observations in which Keskin et al. (8) and Dayanan et al. (19) emphasized that malignancy in GD is predominantly linked to nodular transformation rather than the degree of autoimmune activity alone. These data suggest that, particularly in endemic regions, glandular remodeling and nodule biology represent the principal oncogenic drivers in GD.

Beyond clinical associations, emerging molecular evidence indicates that TRAb-mediated stimulation, chronic inflammation, and oxidative stress may activate pro-carcinogenic signaling pathways within the hyperplastic Graves' thyroid. Vargas-Uricoechea (21) demonstrated that autoimmune hyperactivation is associated with increased DNA damage and enhanced proliferative signaling, supporting a mechanistic link between autoimmune stimulation and carcinogenesis. However, environmental determinants—particularly iodine status—appear to modulate this process. Pellegriti et al. highlighted that chronic iodine deficiency promotes long-standing TSH-driven hyperplasia and nodular transformation, creating a permissive environment for clonal expansion of mutated follicular cells (12). Integration of our cohort findings with these clinical and biological data suggests that non-NGD may represent an autoimmunity-dominant phenotype, whereas NGD reflects chronic structural remodeling within an iodine-deficient milieu and carries a greater carcinogenic propensity. These pathways likely interact rather than act independently, explaining the heterogeneous malignant potential observed across GD phenotypes.

In our cohort, most tumors were papillary microcarcinomas; however, NGD was associated with significantly larger PTCs and higher lymphatic invasion rates than in non-NGD, whereas multifocality, bilaterality, vascular invasion, ETE, and lymph node metastasis did not differ significantly between groups. This pattern echoes prior observations that, although many GD-associated cancers are microcarcinomas, they may still display aggressive features in a subset of patients (5,8,13). Classic reports by Belfiore et al. (22) suggested that PTCs arising in GD are more often multifocal, locally invasive, and nodal or distantly metastatic than cancers in euthyroid patients and advocated for "vigorous" treatment in this subgroup (20). Subsequent analyses have been heterogeneous, with some studies confirming more aggressive behavior and others reporting similar or even more favorable outcomes compared with non-GD controls (12,22,23). A recent systematic review and meta-analysis by Mekraksakit et al. (23) concluded that, although GD is clearly associated with an increased incidence of DTC, its independent impact on prognosis remains uncertain, with most series showing excellent long-term survival (22). Our findings are consistent with this nuanced picture: nodularity appears to mark a subgroup characterized by biologically more active tumors, while the overall profile remains dominated by low-risk PTC.

Our findings also highlight the diagnostic limitations of FNAB in GD, particularly within structurally heterogeneous and hyperfunctioning glands. Although FNAB was performed more frequently in the nodular group, malignancy rates remained high across all Bethesda classes and even among patients without biopsy, with 22-52% harboring PTC. These patterns closely parallel those reported by Wang et al. (24), in whom 53.3% of cytologically benign nodules were malignant at final pathology, emphasizing the high false-negative rate of FNAB in GD due to diffuse hypervascularity, microscopic multifocal disease, and sampling error (24). Importantly, several studies have provided mechanistic explanations for the suboptimal performance of FNAB in patients with GD. Autoimmune and hyperfunctioning glands characteristically exhibit pronounced hypervascularity, coalescent or poorly demarcated nodules, and markedly disrupted parenchymal architecture, all of which hinder accurate needle targeting and reduce specimen cellularity (25,26). These structural alterations contribute to higher false-negative rates and sampling error. This is particularly relevant when microcarcinomas are embedded within heterogeneous or inflamed tissue (27). Sampling from adjacent benign parenchyma rather than the malignant focus has been documented as a common mechanism of misdiagnosis, especially in nodules arising in a mixed autoimmune background (28). In a dedicated Graves cohort, Hang et al. (29) reported that the sensitivity of FNAB for PTC was substantially lower than expected for euthyroid nodular disease, with microcarcinomas among the lesions most frequently underdiagnosed. These findings support the concept that both the biological environment of the thyroid gland and the technical limitations of FNAB synergistically reduce the diagnostic yield. Collectively, emerging evidence, including our results, supports a more liberal FNAB strategy, careful ultrasound surveillance, and the recognition that benign cytology or the absence of FNAB does not reliably exclude PTC in patients with GD. These findings suggest that FNAB may have reduced diagnostic sensitivity in GD due to glandular hypervascularity and heterogeneous parenchymal architecture.

In our study, older age and higher BMI emerged as independent predictors of nodularity, highlighting a structural rather than an immunologic basis for nodule formation in GD. Our results are strongly supported by recent studies in Turkish endemic regions, which report that NGD patients are significantly older and more likely to harbor TC. Kefeli et al. (5) found that more than 50% of GD patients with synchronous nodules in an iodine-deficient area had PTC—many with aggressive features—and that nodular disease clustered in an older demographic. Similarly, Dayanan et al. (19) reported that age and structural thyroid enlargement, rather than autoimmune severity, predicted nodularity and were associated with higher malignancy rates in surgically treated GD patients. These studies closely mirror our findings and emphasize the combined effects of iodine deficiency and structural remodeling on nodule formation. A multicenter study by Yoon et al. (13) (15,159 GD patients; 262 cancers) also supports a structural-age effect: GD patients with coexisting nodules were significantly older (61.9±12.8 vs. 52.1±13.3 years) and more frequently female, while autoimmune activity (TBII levels, GD activity status) showed no difference between nodular and non-nodular cancers. This is consistent with our observation that TSI and anti-TPO were not independent predictors of nodularity. Yoon et al. (13) found that, although NGD patients were older, aggressive PTC variants, such as tall-cell and solid, were more common in the non-NGD group, suggesting that autoimmunity-driven and structural nodular phenotypes represent distinct biological pathways. This dual-phenotype concept of non-nodular, autoimmune-dominant GD versus nodular, structurally remodeled GD is consistent with our data and with several review studies proposing parallel mechanisms of disease expression (10,11,22). Taken together, our findings support the view that age-related glandular remodeling and BMI-associated proliferative signaling are major determinants of nodularity in GD, whereas classical autoimmune markers predict systemic

manifestations (TSI and ophthalmopathy) and do not independently contribute to nodule formation. This distinction is clinically relevant: NGD appears to represent a structurally dominant phenotype—often seen in older or metabolically at-risk individuals—and is associated with a higher prevalence of synchronous PTC, whereas non-NGD reflects an autoimmune-dominant phenotype with higher TSI levels and ophthalmopathy, but not necessarily increased structural disease.

Study Limitations

This study has several limitations. Its retrospective, single-center design and the inclusion of only surgically treated GD patients introduce selection bias, thereby limiting the generalizability to medically managed populations. The study population likely represents a selected higher-risk subgroup enriched for nodularity and/or suspicious clinical–ultrasonographic features; therefore, the observed malignancy rates cannot be generalized to all patients with GD. Ultrasonography was operator-dependent, and variability in nodule characterization or biopsy indications may have affected both nodule detection and FNAB performance. Conducting the study in an iodine-deficient/endemic goiter region may also have contributed to the high malignancy rates observed and may not reflect the situation in iodine-sufficient populations. Additionally, some subgroup analyses—particularly FNAB-based comparisons—were restricted by small sample sizes. Finally, a multivariate model with malignancy as the dependent variable was not constructed because the cohort consisted exclusively of surgically treated patients, representing a selected high-risk population enriched for nodular disease. Therefore, the association between nodularity and malignancy should be interpreted as cohort-specific rather than as an independent predictive model of TC. These limitations should be considered when interpreting the results, and causality cannot be inferred due to the observational design.

Despite these limitations, the study has several important strengths. It represents a well-characterized endocrine surgery cohort from an iodine-deficient region with systematic ultrasonographic evaluation and uniform histopathologic assessment. The direct comparison of nodular and non-NGD within the same clinical setting allowed evaluation of structural versus autoimmune determinants of malignancy risk. Furthermore, the findings provide clinically relevant support for a structural phenotype model of nodular NGD associated with increased TC risk.

Conclusion

In this large surgical cohort from an iodine-deficient region, thyroid malignancy was not uncommon in GD and was particularly pronounced in patients with NGD. These findings support the existence of two clinically relevant GD phenotypes: a nodular, structurally remodeled phenotype associated with an increased risk of cancer and an autoimmune-dominant, non-nodular phenotype. Importantly, malignancy was observed across cytological categories and even in patients without preoperative FNAB, underscoring that neither benign cytology nor the absence of biopsy reliably excludes PTC in GD. In surgically treated GD patients, nodularity may represent a high-risk structural phenotype warranting vigilant US surveillance, a low threshold for FNAB, and careful, individualized decision-making regarding surgical versus medical therapy in GD, especially in endemic goiter areas where cancer prevalence is high; however, these findings should not be extrapolated to non-surgical GD populations without further prospective data.

Ethics

Ethics Committee Approval: This study was approved by the University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital Institutional Ethics Committee (approval no: 463, date: 17.12.2025).

Informed Consent: Due to the retrospective nature of the study and the use of anonymized data, the ethics committee waived the requirement for written informed consent.

Footnotes

Authorship Contributions

Surgical and Medical Practices: T.M.O., M.Y., S.A., A.C., G.Y., S.S., Concept: T.M.O., S.A., G.Y., N.Y.E., S.S., Design: T.M.O., M.Y., A.C., Data Collection or Processing: T.M.O., M.Y., G.Y., N.Y.E., Analysis or Interpretation: M.Y., S.A., A.C., G.Y., N.Y.E., S.S., Literature Search: T.M.O., S.A., A.C., N.Y.E., S.S., Writing: T.M.O.

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

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

References

1. Stan MN, Toro-Tobon D. Graves' disease: a new era of pathophysiology-guided therapeutics. J Clin Endocrinol Metab. 2026:dgag010.
2. Soares MN, Borges-Canha M, Neves C, Neves JS, Carvalho D. The role of Graves' disease in the development of thyroid nodules and thyroid cancer. Eur Thyroid J. 2023;12:e230055.
3. Ross DS, Burch HB, Cooper DS, et al. 2016 American Thyroid Association guidelines for diagnosis and management of hyperthyroidism and other causes of thyrotoxicosis. Thyroid. 2016;26:1343-421.
4. Vural M, Koc E, Evliyaoglu O, et al. Iodine status of Turkish pregnant women and their offspring: a national cross-sectional survey. J Trace Elem Med Biol. 2021;63:126664.
5. Kefeli M, Gucer H, Durmuş ET, Atmaca A, Colak R, Mete O. Prevalence of papillary thyroid carcinoma is significantly higher

in graves disease with synchronous thyroid nodules. Turk Patoloji Derg. 2024;40:190-5.

6. Erbil Y, Barbaros U, Ozbey N, et al. Graves' disease, with and without nodules, and the risk of thyroid carcinoma. J Laryngol Otol. 2008;122:291-5.
7. Kraimps JL, Bouin-Pineau MH, Mathonnet M, et al. Multicentre study of thyroid nodules in patients with Graves' disease. Br J Surg. 2000;87:1111-3.
8. Keskin C, Sahin M, Hasanov R, et al. Frequency of thyroid nodules and thyroid cancer in thyroidectomized patients with Graves' disease. Arch Med Sci. 2019;16:302-07.
9. Papanastasiou A, Sapalidis K, Goulis DG, et al. Thyroid nodules as a risk factor for thyroid cancer in patients with Graves' disease: a systematic review and meta-analysis of observational studies in surgically treated patients. Clin Endocrinol (Oxf). 2019;91:571-7.
10. Huang W, Chen WH. Clinicopathologic predictors of incidental thyroid cancer in Graves' disease: a single-center retrospective cohort study. Updates Surg. 2026;78:319-25.
11. Palella M, Giustolisi FM, Modica Fiascaro A, et al. Risk and prognosis of thyroid cancer in patients with graves' disease: an umbrella review. Cancers (Basel). 2023;15:2724.
12. Pellegriti G, Frasca F, Regalbuto C, Squatrito S, Vigneri R. Worldwide increasing incidence of thyroid cancer: update on epidemiology and risk factors. J Cancer Epidemiol. 2013;2013:965212.
13. Yoon JH, Jin M, Kim M, et al. Clinical characteristics and prognosis of coexisting thyroid cancer in patients with Graves' disease: a retrospective multicenter study. Endocrinol Metab (Seoul). 2021;36:1268-76.
14. Boutzios G, Vasileiadis I, Zapanti E, et al. Higher incidence of tall cell variant of papillary thyroid carcinoma in Graves' disease. Thyroid. 2014;24:347-54.
15. Rotondi M, Coperchini F, Latrofa F, Chiovato L. Role of chemokines in thyroid cancer microenvironment: Is CXCL8 the main player? Front Endocrinol (Lausanne). 2018;9:314.
16. Ali SZ, Baloch ZW, Cochand-Priollet B, Schmitt FC, Vielh P, VanderLaan PA. The 2023 Bethesda System for Reporting Thyroid Cytopathology. Thyroid. 2023;33:1039-44.
17. Baloch ZW, Asa SL, Barletta JA, et al. Overview of the 2022 WHO classification of thyroid neoplasms. Endocr Pathol. 2022;33:27-63.
18. Moronta S, Slattery L, Wang R, Nwariaku F, McMullin JL. Corrigendum to 'incidental thyroid cancer in patients with graves' disease: not as rare as we previously believed' [journal of surgical research volume 308 (2025) 122-128]. J Surg Res. 2025;310:397.
19. Dayanan R, Bilen A, Demirci T, et al. Investigation on the prevalence of thyroid cancer in Graves' patients in northeastern part of Turkey: is surgery a better option for patients with Graves' disease who develop antithyroid drug-related major adverse events? Eur Rev Med Pharmacol Sci. 2022;26:3562-9.
20. Gulsoy Kirnap N, Anil C, Bozkus Y. Graves' disease does not pose an increased risk of thyroid cancer. Annals of Medical Research. 2019;26:2553-6.
21. Vargas-Uricoechea H. autoimmune thyroid disease and differentiated thyroid carcinoma: a review of the mechanisms that explain an intriguing and exciting relationship. World J Oncol. 2024;15:14-27.
22. Belfiore A, Garofalo MR, Giuffrida D, et al. Increased aggressiveness of thyroid cancer in patients with Graves' disease. J Clin Endocrinol Metab. 1990;70:830-5.
23. Mekraksakit P, Rattanawong P, Karnchanasorn R, et al. Prognosis of differentiated thyroid carcinoma in patients with graves disease: a systematic review and meta-analysis. Endocr Pract. 2019;25:1323-37.
24. Wang R, Disharoon M, Song Z, et al. Incidental but not insignificant: thyroid cancer in patients with graves disease. J Am Coll Surg. 2024;238:751-8.
25. Kwon H, Moon BI. Prognosis of papillary thyroid cancer in patients with Graves' disease: a propensity score-matched analysis. World J Surg Oncol. 2020;18:266.
26. Russell MD, Orloff LA. Ultrasonography of the thyroid, parathyroids, and beyond. HNO. 2022;70:333-44.
27. Centeno BA, Szyfelbein WM, Daniels GH, Vickery AL Jr. Fine needle aspiration biopsy of the thyroid gland in patients with prior Graves' disease treated with radioactive iodine. Morphologic findings and potential pitfalls. Acta Cytol. 1996;40:1189-97.
28. Zhu Y, Song Y, Xu G, Fan Z, Ren W. Causes of misdiagnoses by thyroid fine-needle aspiration cytology (FNAC): our experience and a systematic review. Diagn Pathol. 2020;15:1.
29. Hang JF, Lilo MT, Bishop JA, Ali SZ. Diagnostic accuracy of fine needle aspiration in thyroid nodules arising in patients with graves disease. Acta Cytol. 2017;61:117-24.

Original Article

DOI: 10.4274/haseki.galenos.2026.64936
Med Bull Haseki 2026;64(2):138-146

Predictors of Postoperative Sepsis after Percutaneous Nephrostomy for Stone-related Urinary Tract Obstruction: A Tertiary Referral Center Experience

Ramazan Ugur, Emin Taha Keskin, Ahmet Eren Sagir, Abdullah Zilan, Direnc Ozboru, Halil Lutfi Canat

University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital, Clinic of Urology, Istanbul, Türkiye

Abstract

Aim: Percutaneous nephrostomy (PN) is a well-established method for urgent decompression in stone-related urinary obstruction; however, post-procedural sepsis remains a significant concern. This study aimed to determine the incidence and risk factors for sepsis following PN in patients with stone-related urinary tract obstruction who did not have pre-existing sepsis.

Methods: This retrospective observational cohort included patients who underwent PN for stone-related urinary tract obstruction. Patients were classified as non-septic (n=290) or septic (n=18) based on postoperative (post-op) sepsis. Demographic, clinical, laboratory, microbiological, and radiological parameters were analyzed. Univariate and multivariable logistic regression analyses were performed to identify independent predictors of post-op sepsis. Receiver operating characteristic (ROC) curve analysis was performed to evaluate the discriminative performance of the prediction model.

Results: Three hundred and eight patients were included; 18 (5.9%) developed post-op sepsis. Patients with post-op sepsis had significantly higher preoperative (pre-op) body temperature, white blood cell count and neutrophil count, C-reactive protein (CRP) levels, neutrophil-to-lymphocyte ratio, and systemic immune-inflammation index. Bacterial growth in nephrostomy catheter cultures and perirenal fat stranding were more frequent in the sepsis group. Multivariable analysis identified elevated body temperature, higher pre-op neutrophil count, increased pre-op CRP, and bacterial growth in nephrostomy catheter cultures as independent predictors of post-op sepsis. ROC curve analysis showed an area under the curve (AUC) of 0.925 for the prediction model, while body temperature, pre-op CRP, and neutrophil count had AUCs of 0.777, 0.750, and 0.746, respectively.

Conclusion: Percutaneous nephrostomy relieves urinary tract obstruction but may also lead to serious complications such as sepsis. An increased body temperature, a high pre-op neutrophil count, a high pre-op CRP level, and bacterial growth in nephrostomy catheter cultures were all independent predictors, and the prediction model showed excellent discrimination.

Keywords: Percutaneous nephrostomy, sepsis, urinary tract obstruction, risk factors, urolithiasis

Introduction

Drainage of upper urinary tract obstruction caused by benign or malignant conditions is achieved by percutaneous nephrostomy (PN) or double J stent placement (1-3). Percutaneous nephrostomy is frequently performed for urinary diversion in conditions such as urolithiasis, pregnancy, ureteral strictures, obstruction-related sepsis, retroperitoneal fibrosis, pelvic malignancies, and trauma-associated urinary extravasation (4-6). Despite being a commonly performed urological intervention, PN is not free of complications and may be associated with bleeding, urinary tract perforation, urinoma formation due to extravasation, injury to adjacent organs such as the spleen, liver, bowel, or pleura, infection, and nephrocutaneous fistula formation (4,5,7,8).

Stone-related urinary tract obstruction represents a urological emergency requiring prompt decompression

Corresponding Author: Ramazan Ugur, MD, University of Health Sciences Türkiye, Basaksehir Cam and Sakura City Hospital, Clinic of Urology, Istanbul, Türkiye
E-mail: rugur23@gmail.com **ORCID:** orcid.org/0000-0002-0593-8589
Received: 31.01.2026 **Accepted:** 07.03.2026 **Publication Date:** 27.03.2026

Cite this article as: Ugur R, Keskin ET, Sagir AE, Zilan A, Ozboru D, Canat HL. Predictors of postoperative sepsis after percutaneous nephrostomy for stone-related urinary tract obstruction: a tertiary referral center experience. Med Bull Haseki. 2026;64(2):138-146

of the collecting system (4,9). In this situation, PN is frequently preferred due to its rapid effectiveness and wide availability (10). However, despite its effectiveness, manipulation of an infected and obstructed urinary system may lead to bacteremia, endotoxin release, and pyelovenous reflux, potentially resulting in sepsis and even mortality (3,6,11).

We hypothesized that postoperative (post-op) sepsis following PN in patients with stone-related urinary tract obstruction could be anticipated through accessible clinical, laboratory, microbiological, and radiological indicators. Therefore, this study aimed to determine the incidence of post-op sepsis and to identify its independent predictors in patients devoid of pre-existing sepsis.

Materials and Methods

Compliance with Ethical Standards

Ethical approval was obtained from the University of Health Sciences Türkiye Basaksehir Cam and Sakura City Hospital Institutional Ethics Committee at our center (approval no.: E-96317027-514.10-295318989/KAEK/19.11.2025.412, date: 24.11.2025), and the study was conducted in accordance with the Declaration of Helsinki. Prior to the procedures, written informed consent was obtained from all patients.

Study Design

This retrospective observational cohort study was conducted at a tertiary care center and included patients who underwent PN placement for urinary drainage due to stone-related urinary tract obstruction. Patients were excluded if they had evidence of sepsis at the time of PN placement. Patients with malignancy-related external compression, retroperitoneal fibrosis, or bladder, prostate, or gynecological malignancies were also excluded. Additionally, patients with congenital or acquired anatomical strictures, pediatric and pregnant patients, and those with incomplete clinical data were excluded. Patients were divided into two groups based on the development of post-op sepsis: (non-sepsis) and (sepsis) (Figure 1).

Post-operative sepsis was defined according to the Sepsis-3 criteria and recorded if it occurred within 72 hours of PN placement. Sepsis was considered present in patients with suspected or documented infection accompanied by organ dysfunction, defined as an acute increase of ≥2 points in the Sequential Organ Failure Assessment score (12). The primary outcome of the study was the development of post-op sepsis within 72 hours after PN placement.

Demographic, clinical, laboratory, microbiological, and radiological data were obtained from electronic medical records and patient files. The parameters included gender, age, body mass index, diabetes mellitus, history of a solitary kidney, and previous kidney surgery. Vital signs at admission were also recorded, including body temperature, heart rate, respiratory rate, and systolic and diastolic blood pressure. Data on the need for intensive care unit (ICU) admission, bleeding requiring angiographic embolization, and survival status were also included (Table 1). Laboratory

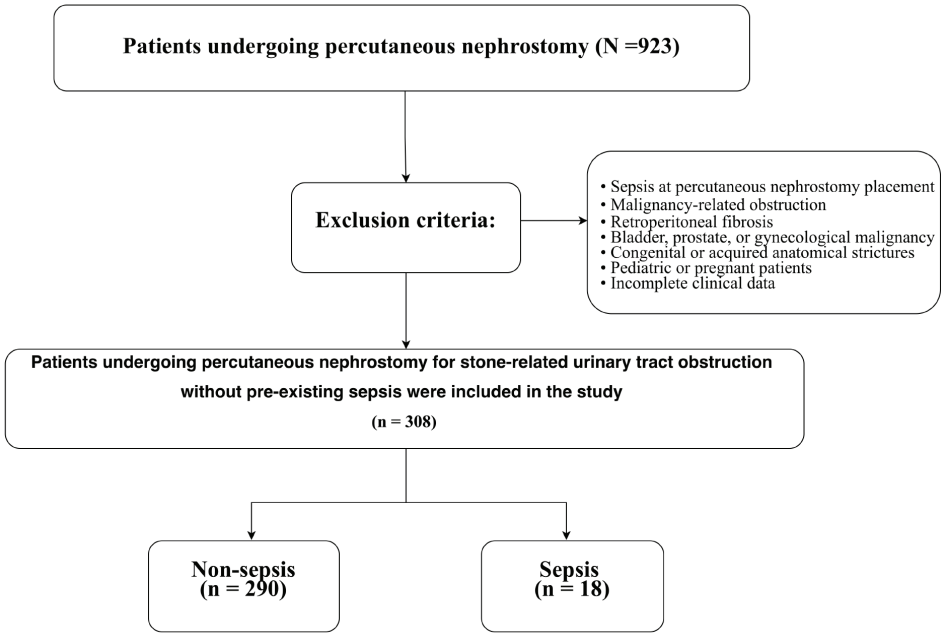

Figure 1. Study flow diagram demonstrating patient selection and group allocation

Table 1. Baseline demographic, clinical, and laboratory characteristics of all patients and stratified by postoperative sepsis status			
	Non-sepsis (n=290)	**Sepsis (n=18)**	**p**
Age, years (mean ± SD)	52.0±17.0	55.8±18.0	0.409†
	n (%)		
Gender			0.240*
Male	156 (53.8)	13 (72.2)	
Female	134 (46.2)	5 (27.8)	
Diabetes mellitus	78 (26.9)	3 (16.7)	0.144*
Solitary kidney	31 (10.7)	4 (22.2)	0.813*
Previous kidney surgery	66 (22.8)	6 (33.3)	0.803*
Urine nitrite positivity	41 (14.1)	6 (33.3)	0.509*
Pyuria	101 (34.8)	15 (83.3)	0.110*
Nephrostomy catheter culture obtained (+)	169 (58.3)	17 (94.4)	**0.011***
Nephrostomy catheter culture growth (+)	59 (20.3)	11 (61.1)	**0.003***
Pre-op urine culture obtained (+)	131 (45.2)	14 (77.8)	0.599*
Pre-op urine culture growth (+)	62 (21.4)	6 (33.3)	0.943*
Post-op ICU requirement	2 (0.7)	17 (94.4)	**<0.001***
Post-op bleeding requiring AE	8 (2.8)	6 (33.3)	0.324*
Mortality	8 (2.8)	3 (16.7)	0.601*
	Median (IQR)		
Body mass index	27.6 (8.4)	24.10 (6.8)	0.234#
Body temperature, °C	36.5 (0.3)	37.05 (1.4)	**<0.001#**
Heart rate, bpm	82 (16)	85 (30)	0.293#
Respiratory rate, /min	20 (3)	20 (4)	0.129#
Systolic blood pressure, mmHg	125 (25)	120 (28)	0.401#
Diastolic blood pressure, mmHg	74 (14)	74.5 (13)	0.721#
Pre-op creatinine, mg/dL	1.35 (1.73)	2.05 (4.69)	0.097#
Pre-op WBC count, ×10³/µL	9.53 (5.08)	15.00 (9.29)	**<0.001#**
Pre-op neutrophil count, ×10³/µL	6.9 (4.96)	11.2 (5.74)	**<0.001#**
Pre-op platelet count, ×10³/µL	272.5 (124)	259.0 (201)	0.512#
Pre-op lymphocyte count, ×10³/µL	1.90 (1.11)	1.47 (1.35)	**0.027#**
Pre-op CRP, mg/L	50 (112)	187 (181)	**<0.001#**
Pre-op procalcitonin, ng/mL	0.20 (1.54)	0.61 (5.92)	0.075#
Pre-op arterial blood gas pH	7.32 (0.17)	7.32 (0.08)	0.871#
ALT, U/L	16 (14)	17 (9)	0.522#
AST, U/L	19 (11.5)	25 (18)	**0.028#**
GGT, U/L	18.5 (26.75)	37 (21)	0.317#
Amylase, U/L	53 (50.75)	41 (52.5)	0.131#
Lipase, U/L	24.5 (26)	22.0 (22.5)	0.288#
Pre-op NLR	4.19 (5.45)	10.26 (10.10)	**<0.001#**
Pre-op PLR	175 (114)	208 (144)	0.249#
Pre-op SII	1247 (1465)	2271 (2395)	**0.010#**
Post-op NLR	3.24 (4.08)	4.08 (8.31)	0.110#
Post-op PLR	154 (117)	152 (102)	0.781#
Post-op SII	935 (1133)	1136 (1359)	0.311#
Post-op creatinine, mg/dL	1.21 (1.15)	1.45 (2.19)	0.487#
Post-op CRP, mg/L	46 (89)	104 (159)	**0.005#**
Post-op procalcitonin, ng/mL	0.14 (0.41)	0.74 (8.74)	**<0.001#**
Post-op WBC count, ×10³/µL	8.85 (3.73)	10.20 (5.61)	0.289#
Post-op platelet count, ×10³/µL	269.0 (124)	254 (183)	0.333#
Post-op neutrophil count, ×10³/µL	6.00 (3.88)	7.11 (5.46)	0.306#
Post-op lymphocyte count, ×10³/µL	1.90 (1.11)	1.47 (1.35)	0.162#
Post-op arterial blood gas pH	7.35 (0.13)	7.40 (0.08)	0.120#

†Independent Samples t-test, #Mann-Whitney U test; *Pearson chi-square or Fisher's exact test

AE: Angioembolization, ALT: Alanine aminotransferase, AST: Aspartate aminotransferase, GGT: Gamma-glutamyl transferase, CRP: C-reactive protein, ICU: Intensive care unit, NLR: Neutrophil-to-lymphocyte ratio, PLR: Platelet-to-lymphocyte ratio, Post-op: Postoperative, Pre-op: Preoperative, SD: Standard deviation, SII: Systemic immune-inflammation index, WBC: White blood cell count, IQR: Interquartile range

parameters consisted of serum creatinine, white blood cell (WBC), neutrophil, lymphocyte, and platelet counts, C-reactive protein (CRP), and procalcitonin. Preoperative (pre-op) urine nitrite positivity and pyuria, arterial blood gas pH, alanine aminotransferase, aspartate aminotransferase (AST), gamma-glutamyl transferase, amylase, and lipase were also recorded.

Postoperative laboratory values were recorded from the initial measurements obtained after nephrostomy placement. Inflammatory indices, including the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio, and systemic immune-inflammation index (SII), were evaluated using both pre-op and post-op laboratory values. Midstream urine cultures were obtained before PN placement, and additional urine samples were collected from the nephrostomy catheter at the time of insertion. Empirical intravenous third-generation cephalosporin therapy was initiated perioperatively in all patients and subsequently adjusted according to culture and antimicrobial susceptibility results.

Radiological findings were assessed using pre-op computed tomography scans. Evaluated parameters included stone location, perirenal fat stranding, laterality, and hydronephrosis grade. The laterality of nephrostomy placement was also recorded.

Statistical Analysis

Statistical analysis was performed using SPSS version 27.0 software (IBM Corp., Armonk, NY, USA). Categorical variables were summarized as frequencies and percentages, while continuous variables were presented as mean ± standard deviation for normally distributed data or as median and interquartile range for non-normally distributed data. The normality of the distribution of continuous variables was assessed using the Kolmogorov-Smirnov and Shapiro-Wilk tests, along with visual methods (histograms and Q-Q plots) and analytical methods (skewness and kurtosis values). For comparisons between two independent groups, the Independent Samples t-test was used when the data were normally distributed, while the Mann-Whitney U test was applied for non-normally distributed data. Associations between categorical variables were analyzed using the Pearson chi-square or Fisher's exact test as appropriate. To identify factors associated with post-op sepsis, univariate logistic regression analysis was first performed. Variables with statistical significance in univariate analysis were subsequently included in a multivariable logistic regression analysis to determine independent predictors of post-op sepsis. Variables with a p-value <0.05 in univariate analysis were included in the multivariable analysis. Receiver operating characteristic (ROC) curve analysis was performed to evaluate the predictive performance of the developed model and selected pre-op parameters for post-op sepsis following PN. Discriminative ability was assessed using the area under the curve (AUC). Optimal cutoff values were determined using the Youden index derived from the ROC curve. The goodness-of-fit of the model was evaluated using the Hosmer-Lemeshow test. For all statistical analyses, a p-value <0.05 was considered statistically significant.

Results

Among 923 patients assessed for eligibility, 308 met the inclusion criteria and were included in the final analysis. Of these, 290 patients (94.1%) were classified into the non-sepsis group and 18 (5.9%) into the sepsis group. The cohort consisted of 169 male (54.9%) and 139 female (45.1%) patients, and sex distribution did not differ significantly between the groups. Mean age was also similar between the non-sepsis and sepsis groups (52.0±17.0 vs. 55.8±18.0 years, p=0.409) (Table 1).

Patients in Group 2 (sepsis) presented with a significantly higher body temperature than patients in Group 1 (non-sepsis) (37.05 °C vs. 36.5 °C, p<0.001; Table 1).

Compared with the non-sepsis group, patients who developed sepsis had significantly higher pre-op WBC counts, neutrophil counts, CRP levels, NLR, and SII values and lower lymphocyte counts (all p<0.05). Postoperatively, CRP and procalcitonin levels were also significantly higher in the sepsis group. In microbiological analyses, both the rate of nephrostomy catheter culture sampling and the rate of bacterial growth in nephrostomy catheter cultures were significantly higher among patients with sepsis (all p<0.05) (Table 1).

Radiological assessments indicated that perirenal fat stranding was significantly more prevalent in Group 2 (sepsis) (94.4% vs. 39%, p=0.011). No significant differences were observed between the groups in stone location, stone laterality, grade of hydronephrosis, or laterality of nephrostomy tube placement (all p>0.05, Table 2).

The requirement for post-op ICU was significantly more frequent in sepsis group (94.4% vs. 0.7%, p<0.001). Higher rates of post-op bleeding requiring angiographic embolization and in-hospital mortality were observed in sepsis group; these differences did not reach statistical significance (Table 1).

After univariate and multivariable logistic regression analyses, elevated body temperature, bacterial growth in nephrostomy catheter cultures, higher pre-op CRP, and higher pre-op neutrophil count were identified as independent predictors of post-op sepsis (Table 3).

ROC curve analysis demonstrated that the prediction model had an AUC of 0.925. The AUC values for individual

Table 2. Radiological stone characteristics of all patients and stratified by postoperative sepsis status			
	Non-sepsis (n=290)	**Sepsis (n=18)**	**p**
	n (%)		
Perirenal fat stranding	113 (39.0)	17 (94.4)	**0.011***
Stone location			0.501*
Calyceal stone	18 (6.2)	2 (11.1)	
Renal pelvis	45 (15.5)	1 (5.6)	
Ureteropelvic junction	60 (20.7)	6 (33.3)	
Proximal ureter	91 (31.4)	5 (27.8)	
Mid-ureter	30 (10.3)	2 (11.1)	
Distal ureter	46 (15.9)	2 (11.1)	
Hydronephrosis			0.803*
None	6 (2.1)	0 (0.0)	
Unilateral	224 (77.2)	13 (72.2)	
Bilateral	60 (20.7)	5 (27.8)	
Hydronephrosis, grade			0.097*
0	14 (4.8)	1 (5.6)	
1	24 (8.3)	1 (5.6)	
2	138 (47.6)	5 (27.8)	
3	81 (27.9)	10 (55.6)	
4	33 (11.4)	1 (5.6)	
Nephrostomy			0.794*
Unilateral	224 (77.2)	14 (77.8)	
Bilateral	66 (22.8)	4 (22.2)	
*Pearson chi-square or Fisher's exact test			

Table 3. Univariate and multivariable logistic regression analysis to determine prognostic factors for post-op sepsis after nephrostomy tube placement								
	Univariate analysis				**Multivariable analysis**			
	OR	**95% CI**		**p-value**	**OR**	**95% CI**		**p-value**
		Lower	**Upper**			**Lower**	**Upper**	
Perirenal fat stranding	3.528	1.270	9.800	**0.016**	0.889	0.143	5.534	0.900
Body temperature	7.275	3.020	17.524	**<0.001**	15.174	3.119	73.836	**<0.001**
Nephrostomy catheter culture (+)	5.766	1.289	25.784	**0.022**	5.879	0.182	189.631	0.318
Nephrostomy catheter culture growth	4.384	1.555	12.354	**0.005**	13.648	2.205	84.454	**0.005**
Pre-op WBC count	1.168	1.070	1.274	**<0.001**	1.115	0.936	1.329	0.224
Pre-op neutrophil count	1.124	1.040	1.214	**0.003**	1.362	1.075	1.724	**0.010**
Pre-op lymphocyte count	0.546	0.265	1.125	0.101	-	-	-	-
AST	1.006	0.988	1.025	0.509	-	-	-	**-**
Pre-op NLR	1.125	1.043	1.214	**0.002**	0.676	0.277	1.649	0.390
Pre-op CRP	1.009	1.004	1.014	**<0.001**	1.014	1.004	1.023	**0.004**
Pre-op SII	1.002	0.995	1.012	**0.003**	1.000	0.997	1.003	0.719
Post-op CRP	1.003	0.996	1.009	0.728	-	-	-	**-**
Post-op procalcitonin	1.001	0.998	1.002	**0.027**	1.000	0.998	1.004	0.515
CI: Confidence interval, AST: Aspartate aminotransferase, CRP: C-reactive protein, NLR: Neutrophil-to-lymphocyte ratio, OR: Odds ratio, PLR: Platelet-to-lymphocyte ratio, Post-op: Postoperative, Pre-op: Preoperative, SII: Systemic immune-inflammation index, WBC: White blood cell count								

parameters were 0.777 for body temperature, 0.750 for pre-op CRP, and 0.746 for pre-op neutrophil count (Figure 2). Based on the Youden index, the optimal cut-off values were 37.05 °C for body temperature (sensitivity 50.0%, specificity 94.7%), 98.5 mg/L for pre-op CRP (sensitivity 77.8%, specificity 67.1%), and 9.41×10³/μL for pre-op neutrophil count (sensitivity 77.8%, specificity 72.5%). The Hosmer-Lemeshow goodness-of-fit test indicated an adequate model fit (p=0.443).

Discussion

In cases of upper urinary tract obstruction, emergency drainage is recommended (2,13). It has been reported that there is no definitive superiority between double J stenting and PN (14). The choice of drainage method has a decisive influence not only on providing mechanical decompression but also on systemic inflammatory response, sepsis, and procedure-related complications (4,10). While PN is an effective and rapid drainage method, it can lead to systemic and local complications, particularly in infected systems, due to the direct involvement of the renal parenchyma and pelvicalyceal system (3,4,15). From a pathophysiological perspective, this intervention, by resembling a high-grade (grade IV) renal injury, may facilitate the translocation of infected urine into the systemic circulation via pyelovenous and pyelolymphatic pathways (16,17). Considering that the kidneys receive approximately 20-25% of the cardiac output, have a blood flow of about 1.0-1.2 L per minute, and have a daily glomerular filtration volume of approximately 180 L, infected content may rapidly enter the systemic circulation via pyelovenous and pyelolymphatic pathways, potentially contributing to the development of a systemic inflammatory response (18-20). In our study, the identification of growth in nephrostomy catheter cultures as an independent predictor of post-op sepsis is consistent with the concept that PN may facilitate systemic dissemination of infected urinary contents through pyelovenous and pyelolymphatic pathways. Conversely, the absence of a significant association between post-op sepsis and pre-op urinalysis findings (pyuria and nitrite positivity), as well as a positive urine culture, suggests that sepsis risk may be more closely related to infected material in the upper urinary tract. Accordingly, the selection of the drainage method should be individualized based on the patient's clinical condition, anatomical characteristics, hemodynamic stability, and the technical feasibility of the procedure.

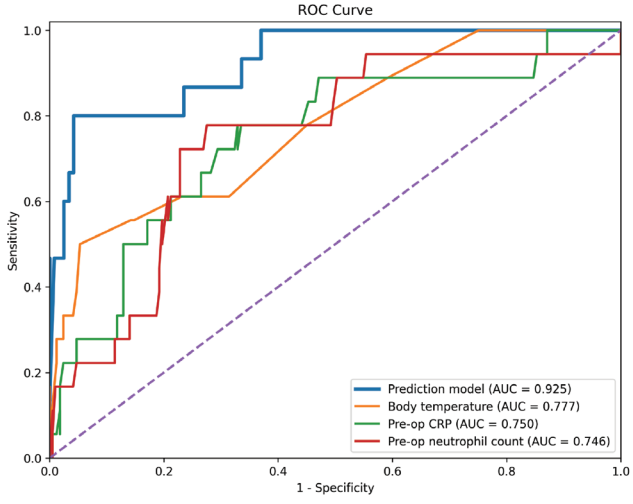

Figure 2. ROC curve analysis for the prediction of postoperative sepsis following percutaneous nephrostomy

ROC: Receiver operating characteristic, AUC: Area under the curve, Pre-op: Preoperative, CRP: C-reactive protein

Elevated body temperature, increased WBC count—particularly the neutrophil count—and high CRP levels are early, readily available indicators of active systemic inflammation, and their association with sepsis risk has been reported previously (21-23). Fever is a systemic response that occurs as a result of the effects of pyrogenic cytokines (interleukin-1, tumor necrosis factor-α, interleukin-6, and interferons) on the hypothalamic thermoregulatory center and represents an early indicator of systemic inflammatory activation (24). In the presence of an obstructed upper urinary system, elevated body temperature, which was identified as an independent predictor in our study, should be regarded as a clinical warning sign indicating potential development of sepsis. C-reactive protein, an acute-phase reactant, is stimulated in the liver via the interleukin-6 pathway and is a sensitive biomarker reflecting the severity of infection and the magnitude of the inflammatory response (25). In the literature, high CRP levels have been reported to be associated with bacteremia and sepsis (9). Similarly, neutrophils are key effector cells of the immune response in sepsis, with increased neutrophil counts reflecting an early and intense inflammatory response to infection. However, it has been shown that excessive neutrophil activation and associated cytotoxic mediator release can contribute to sepsis pathogenesis through tissue damage and microvascular dysfunction (26). The development of post-op sepsis is associated with increased pre-op CRP levels and pre-op neutrophil counts, which can be considered a pathophysiological reflection of the systemic inflammatory response in obstructed and infected systems.

In obstructed upper urinary tracts, the need for ICU after PN is a crucial indicator reflecting the clinical severity of sepsis and is associated with increased morbidity, prolonged hospital stay, and increased healthcare costs (3,10,27). Sepsis developed in 89% (17/19) of patients who required the ICU after PN in our study, supporting the clinical significance of this relationship.

Radiologically, the presence of perirenal fat stranding on computed tomography represents a marker of local inflammation and, in our study, was found to be significantly associated with post-op sepsis (28). However, the lack of independent predictive value of this variable in the multivariable logistic regression analysis suggests that local inflammation may increase the risk of sepsis in the presence of a concomitant systemic inflammatory response. Similarly, pre-op lymphocyte count, AST, pre-op NLR, pre-op SII, post-op CRP, and post-op procalcitonin were associated with sepsis in univariate analyses but were not independent predictors in the multivariable analysis. This finding indicates that these parameters may reflect indirect manifestations of systemic inflammation and overall disease severity rather than serving as stand-alone determinants of post-op sepsis risk, underscoring the need to interpret them within a comprehensive clinical context. From a clinical perspective, all these variables may help identify patients who require closer monitoring, earlier escalation of antimicrobial therapy, and heightened post-procedural surveillance.

Percutaneous nephrostomy is an invasive procedure that may be associated with a broad spectrum of complications, including bleeding and various local adverse events. In particular, the close anatomical proximity of structures such as the colon, liver, spleen, pleura, and lungs may increase the risk of injury to these organs, which, in turn, may adversely affect the clinical course. In addition, urinary extravasation and catheter-related problems can lead to limitations in daily activities and reduced quality of life, particularly for patients who may experience chronic pain or recurrent infections as a result of these complications (7,29-31). In our study, bleeding requiring angioembolization was identified in 14 patients (4.5%). This complication is typically attributable to procedure-related segmental arterial injury, arteriovenous fistula formation, or pseudoaneurysm development (29,32). Notably, in patients with coagulopathy and radiological evidence of perirenal fat stranding, increased tissue fragility is likely to increase the potential risks associated with PN (33). Therefore, procedure-related complications should be carefully considered when selecting a drainage strategy. In patients in whom retrograde access is feasible, retrograde double J stent placement may represent a less invasive option without direct trauma to the renal parenchyma; conversely, PN may be preferred when retrograde access is not feasible or has failed.

Study Limitations

This study has several limitations. The retrospective design and the absence of a comparative control group evaluating PN limit the ability to draw method-specific inferences from the observed findings. In addition, ICU requirement, which represents a clinical outcome resulting from sepsis, was not included in the logistic regression analysis to minimize potential bias and allow a more appropriate assessment of causal relationships. Another limitation is the relatively low number of sepsis events (n=18). Since only 18 instances of sepsis were recorded, the multivariable model may be susceptible to overfitting; consequently, the identified predictors should be regarded as preliminary until externally validated. Finally, the single-center nature of the study represents an additional limitation that may restrict the generalizability of the findings. Furthermore, the predictive model lacked internal bootstrapping and external validation; consequently, its discriminative efficacy may be overly optimistic and requires validation in independent cohorts.

Despite these limitations, this study includes a relatively large, well-characterized cohort of patients undergoing PN for stone-related urinary tract obstruction without pre-existing sepsis. Furthermore, it identifies independent predictors of post-op sepsis, allowing identification of higher-risk patients in this population.

Conclusion

This study underscores the potential prognostic significance of clinical and laboratory parameters in forecasting sepsis subsequent to PN in patients with obstructed upper urinary tracts. An elevated body temperature, higher pre-op neutrophil counts and pre-op CRP levels, and bacterial growth in nephrostomy catheter cultures were identified as independent predictors of post-op sepsis. These findings may help clinicians identify patients at increased risk of post-op sepsis and support more individualized peri-procedural monitoring and management. In high-risk patients, it may be clinically appropriate to consider less invasive drainage strategies before proceeding to more invasive approaches. Further prospective randomized controlled studies are needed to clarify the effects of different drainage strategies on sepsis development and clinical outcomes.

Ethics

Ethics Committee Approval: The study was approved by the University of Health Sciences Türkiye Basaksehir Cam and Sakura City Hospital institutional Ethics Committee at our center (approval no.: E-96317027-514.10-295318989/KAEK/19.11.2025.412, date: 24.11.2025).

Informed Consent: Informed consent was obtained from all patients involved in this study.

Footnotes

Authorship Contributions

Surgical and Medical Practices: R.U., E.T.K., A.E.S., A.Z., D.O., H.L.C., Concept: R.U., E.T.K., A.E.S.,A.Z., D.O.,

H.L.C., Design: R.U., E.T.K., A.E.S.,A.Z., D.O., H.L.C., Data Collection or Processing: R.U., E.T.K., A.E.S., A.Z., D.O., H.L.C., Analysis or Interpretation: R.U., E.T.K., A.E.S., A.Z., D.O., H.L.C., Literature Search: R.U., E.T.K., A.E.S., A.Z., D.O., H.L.C., Writing: R.U., E.T.K., A.E.S., A.Z., D.O., H.L.C.

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

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

References

1. Wong R, Ziegler J, Bal DS, Lother SA, Patel P, Rush B. Nephrostomy tube versus ureteral stent for obstructing septic calculi: a nationwide propensity score-matched analysis. J Assoc Med Microbiol Infect Dis Can. 2024;9:73-81.
2. Moon YJ, Jun DY, Jeong JY, Cho S, Lee JY, Jung HD. percutaneous nephrostomy versus ureteral stent for severe urinary tract infection with obstructive urolithiasis: a systematic review and meta-analysis. Medicina (Kaunas). 2024;60:861.
3. Elmansy H, Berjaoui MB, Savage DW, et al. Comparing outcomes of ureteric stent versus nephrostomy tube in urosepsis: a population-based study in Ontario, Canada. World J Urol. 2025;43:692.
4. Yoo MJ, Bridwell RE, Inman BL, Henderson JD, Long B. Approach to nephrostomy tubes in the emergency department. Am J Emerg Med. 2021;50:592-6.
5. Parmar K, Htun HZ, John A, et al. Outcome analysis of percutaneous nephrostomy insertion to relieve urinary obstruction in advanced malignancy. Urol Ann. 2025;17:239-44.
6. Zhang Z, Wang X, Chen D, et al. Minimally invasive management of acute ureteral obstruction and severe infection caused by upper urinary tract calculi. J Xray Sci Technol. 2020;28:125-35.
7. Said DF, Ferreira DB, Hayek KKRE, et al. Factors associated with complications after percutaneous nephrolithotomy: an analysis of 1,066 cases. Rev Assoc Med Bras (1992). 2023;69:e20221089.
8. Paredes-Bhushan V, Raffin EP, Denstedt JD, et al. Outcomes of conservative management of splenic injury incurred during percutaneous nephrolithotomy. J Endourol. 2020;34:811-5.
9. Anıl H, Şener NC, Karamık K, et al. Comparison of percutaneous nephrostomy and ureteral dj stent in patients with obstructive pyelonephritis: a retrospective cohort study. J Invest Surg. 2022;35:1445-50.
10. Lu X, Zhou B, Hu D, Ding Y. Emergency decompression for patients with ureteral stones and SIRS: a prospective randomized clinical study. Ann Med. 2023;55:965-72.
11. Wu H, Fu Y, He Q, Bao J. Risk factors for urosepsis after percutaneous nephrolithotomy: a meta-analysis. Asian J Surg. 2023;46:2408-9.
12. Singer M, Deutschman CS, Seymour CW, et al. The third international consensus definitions for sepsis and septic shock (sepsis-3). JAMA. 2016;315:801-10.
13. Ibrahim RM, Lotfy AM, Mahmoud O, Elrady MA, Abdallah M. Effect of initial drainage method on retrograde intrarenal surgery outcomes in acute calcular pyelonephritis: a prospective comparative study. Urolithiasis. 2026;54:40.
14. Skolarikos A, Jung H, Neisius A, et al. EAU Guidelines on Urolithiasis. 2025. Available from: https://d56bochluxqnz.cloudfront.net/documents/full-guideline/EAU-Guidelines-on-Urolithiasis-2025.pdf Accessed January 20, 2026.
15. Rizzo M, Ongaro L, Claps F, et al. An observational cohort study investigating the incidence rate of infectious complications after routine office nephrostomy tube replacement without using antimicrobial prophylaxis. Urology. 2021;147:27-32.
16. Farag M, Timm B, Davis N, Wong LM, Bolton DM, Jack GS. Pressurized-bag irrigation versus hand-operated irrigation pumps during ureteroscopic laser lithotripsy: comparison of infectious complications. J Endourol. 2020;34:914-8.
17. Hong A, du Plessis J, Browne C, Jack G, Bolton D. Mechanism of urosepsis: relationship between intrarenal pressures and pyelovenous backflow. BJU Int. 2023;132:512-9.
18. Maiti K, Maity S, Santra A, Sarkar D. A prospective study on the outcome of pyonephrosis and infected hydronephrosis drained by percutaneous nephrostomy - a tertiary care centre experience. Turk J Surg. 2025;41:411-7.
19. Kaufman DP, Basit H, Knohl SJ. Physiology, glomerular filtration rate. 2023 Jul 17. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2026 Jan-.
20. Hanessian K, Albaghli A, Crew R, et al. The effect of negative, single and multi-organism positive cultures on outcomes following PCNL. BJUI Compass. 2026;7:e70150.
21. Liu L, Xiao Y, Yue X, Wang Q. Safety and efficacy of enhanced recovery after surgery among patients undergoing percutaneous nephrolithotomy: a systematic review and meta-analysis. Int J Surg. 2024;110:3768-77.
22. Akdeniz E, Ozturk K, Ulu MB, Gur M, Caliskan ST, Sehmen E. Risk factors for systemic inflammatory response syndrome in patients with negative preoperative urine culture after percutaneous nephrolithotomy. J Coll Physicians Surg Pak. 2021;31:410-6.
23. Barros PL, Oliveira B, Pereira R, Dores M, Coutinho A, Peyroteo I. A possible novel approach to small iatrogenic colonic perforation during nephrostomy placement. Urol Case Rep. 2023;50:102495.
24. Luheshi GN. Cytokines and fever. Mechanisms and sites of action. Ann N Y Acad Sci. 1998;856:83-9.
25. Plebani M. Why C-reactive protein is one of the most requested tests in clinical laboratories? Clin Chem Lab Med. 2023;61:1540-5.
26. Kolaczkowska E, Kubes P. Neutrophil recruitment and function in health and inflammation. Nat Rev Immunol. 2013;13:159-75.
27. Pietropaolo A, Hendry J, Kyriakides R, et al. Outcomes of elective ureteroscopy for ureteric stones in patients with prior urosepsis and emergency drainage: prospective study over 5 yr from a tertiary endourology centre. Eur Urol Focus. 2020;6:151-6.
28. Lai C, Hu Z, Zhu J, et al. Development and validation of a deep learning-based automated computed tomography

image segmentation and diagnostic model for infectious hydronephrosis: a retrospective multicentre cohort study. EClinicalMedicine. 2025;82:103146.

29. Alabat Roca A, Torrecilla Ortíz C, Cuadrado Campaña JM, et al. Hemorrhagic complicationes after percutaneous nephrolithotomy: the importance of an early endovascular management. Actas Urol Esp (Engl Ed). 2021;45:635-41.
30. Degirmenci T, Gunlusoy B, Kozacioglu Z, et al. Utilization of a modified Clavien Classification System in reporting complications after ultrasound-guided percutaneous nephrostomy tube placement: comparison to standard Society of Interventional Radiology practice guidelines. Urology. 2013;81:1161-7.
31. Gauhar V, Pirola GM, Scarcella S, et al. Nephrostomy tube versus double J ureteral stent in patients with malignant ureteric obstruction. A systematic review and meta-analysis of comparative studies. Int Braz J Urol. 2022;48:903-14.
32. Zeng T, Wu W, Zhang L, et al. Superselective renal arterial embolization for severe postpercutaneous nephrolithotomy haemorrhage: clinical characteristics and risk factors for initial failure. World J Urol. 202;41:1647-52.
33. Kumar S, Dutt UK, Singh S, et al. Prospective audit of complications after ultrasonography-guided percutaneous nephrostomy for upper urinary tract obstruction using modified Clavien classification system. Urol Ann. 2020;12:31-6.

Original Article

DOI: 10.4274/haseki.galenos.2026.60352
Med Bull Haseki 2026;64(2):147-154

Association Between Sarcopenia Risk, Disease Severity, and Functional Mobility in Parkinson's Disease: A Cross-Sectional Study

Hatice Kubra Asik,, Mert Zure, Dilek Un Oguzhanasilturk, Zeynep Karakuzu Gungor

University of Health Sciences Türkiye, Istanbul Kanuni Sultan Suleyman Training and Research Hospital, Clinic of Physical Medicine and Rehabilitation, Istanbul, Türkiye

Abstract

Aim: Sarcopenia and Parkinson's disease (PD) share pathophysiological mechanisms that may synergistically accelerate functional decline. This study aimed to investigate the relationship between sarcopenia risk and functional mobility in patients with PD and to compare the clinical and demographic characteristics of patients with low and high sarcopenia risk.

Methods: This cross-sectional study enrolled 29 patients with idiopathic PD between October and December 2025. Sarcopenia risk was assessed using the SARC-F questionnaire (cut-off ≥4) and functional mobility using the Timed Up and Go (TUG) test. Disease severity was staged using the Hoehn-Yahr (H-Y) scale. Spearman rank correlation, the Mann-Whitney U test, and independent t-tests were applied; normality was evaluated using the Kolmogorov-Smirnov test.

Results: Of 29 participants (mean age 70.59±12.61 years), 65.5% were classified as high sarcopenia risk. The high-risk group was significantly older (75.42 vs. 63.00 years; p=0.008) and had more advanced H-Y staging (2.37 vs. 1.50; p=0.005). Timed Up and Go time was significantly longer in the high-risk group (27.58 vs. 18.30 seconds; p=0.018). SARC-F score correlated strongly with TUG time (r=0.709, p<0.001) and moderately with H-Y stage (r=0.440, p=0.016).

Conclusion: Sarcopenia risk is highly prevalent in PD and is strongly associated with disease severity and impaired functional mobility. Combined SARC-F and TUG screening may facilitate early identification of at-risk patients to guide targeted rehabilitation interventions.

Keywords: Sarcopenia, Parkinson disease, muscle weakness, postural balance, functional status, geriatric assessment

Introduction

Parkinson's disease (PD) is a chronic, progressive neurodegenerative disorder and the second most prevalent neurodegenerative condition globally, impacting approximately 8.5 million individuals worldwide (1,2). It is marked by the loss of dopaminergic neurons in the substantia nigra pars compacta, which leads to major motor symptoms like rigidity, bradykinesia, postural instability, and resting tremor, as well as major non-motor symptoms (1). Sarcopenia is a progressive and generalized skeletal muscle disorder characterized by diminished muscle strength, reduced muscle mass or quality, and compromised physical performance. It was officially recognized as an independent disease entity (ICD-10 M62.84) and subsequently updated by the European Working Group on Sarcopenia in Older People (EWGSOP2) in 2019 (3). The pathophysiological mechanisms underlying both conditions—such as aging, chronic inflammation, oxidative stress, hormonal changes, and physical inactivity—exhibit significant overlap, indicating a synergistic relationship that may expedite functional decline in affected individuals (4).

The prevalence of sarcopenia in patients with PD varies significantly across studies, ranging from 10.9% to 55.8%, contingent upon diagnostic criteria and assessment

Corresponding Author: Hatice Kubra Asik, MD, University of Health Sciences Türkiye, Istanbul Kanuni Sultan Suleyman Training and Research Hospital, Clinic of Physical Medicine and Rehabilitation, Istanbul, Türkiye
E-mail: drkubraftr@gmail.com **ORCID:** orcid.org/0000-0001-9462-9990
Received: 02.01.2026 **Accepted:** 09.03.2026 **Publication Date:** 27.03.2026

Cite this article as: Asik, HK, Zure M, Un Oguzhanasilturk D, Karakuzu Gungor Z. Association between sarcopenia risk, disease severity, and functional mobility in Parkinson's disease: a cross-sectional study. Med Bull Haseki. 2026;64(2):147-154

methodologies (5-7). Meta-analytic data demonstrate that the prevalence of sarcopenia in PD is roughly threefold greater than that in age-matched controls, with aggregated estimates around 29% (4,8). In addition to decreased muscle mass, sarcopenia in PD is closely linked to a higher frequency of falls, functional disability, cognitive decline, reduced quality of life, and an increased risk of mortality (8,9). Recent rehabilitation research shows that exercise-based treatments, especially progressive resistance training combined with intensive programs from different fields, can significantly improve muscle strength, physical performance, and functional outcomes in patients with PD (10,11).

Despite the growing international literature addressing sarcopenia in patients with PD, comprehensive studies evaluating the interrelationship between sarcopenia risk, functional mobility, and disease severity in Turkish patient populations remain scarce. Early identification of individuals at risk of sarcopenia using validated screening instruments, such as the SARC-F questionnaire, in conjunction with objective functional mobility assessments, including the Timed Up and Go (TUG) test, may provide valuable clinical insights for the development of individualized rehabilitation strategies. Therefore, the present study aimed to investigate the association between sarcopenia risk and functional performance in patients with PD, to compare clinical and demographic characteristics between low- and high-sarcopenia-risk groups, and to explore the correlations among sarcopenia risk scores, disease severity, and functional mobility. We hypothesized that higher sarcopenia risk, as assessed by the SARC-F, would be significantly associated with impaired functional mobility and greater disease severity in patients with PD.

Materials and Methods

Compliance with Ethical Standards

This cross-sectional observational study was performed at the Physical Medicine and Rehabilitation Department of a tertiary training and research hospital between October and December 2025. The University of Health Sciences Türkiye, Istanbul Kanuni Sultan Suleyman Training and Research Hospital Institutional Clinical Research Ethics Committee examined and approved the study protocol (approval number: KAEK/2025.09.235, date: 25.09.2025). All procedures were conducted in compliance with the ethical principles of the Declaration of Helsinki and its subsequent revisions. Written informed consent was obtained from participants after comprehensively outlining the study's goal, procedures, potential benefits, and risks. Data privacy and participant anonymity were carefully preserved throughout the research process. The authors declare no conflict of interest and that no external funding was received for this study.

Study Design and Participants

A cross-sectional design was used to investigate the association between sarcopenia risk and functional status among individuals with PD. Participants were recruited consecutively from patients attending the outpatient clinic. The study flowchart is presented in Figure 1.

Inclusion criteria were (1) clinical diagnosis of PD according to Movement Disorder Society criteria; (2) age 40-80 years; (3) Hoehn-Yahr (H-Y) disease stage 1-4; (4) ability to stand and walk independently, with or without assistive devices; (5) stable medication regimen for at least four weeks before assessment; and (6) cognitive capacity to understand and follow test instructions.

Exclusion criteria included the following: (1) secondary or atypical parkinsonism; (2) severe cognitive impairment that prevents test comprehension; (3) acute medical conditions or infections; (4) severe cardiovascular, respiratory, or orthopedic comorbidities that limit test participation; (5) recent surgery or hospitalization within the past three months; and (6) presence of deep brain stimulation devices.

Demographic data, including age, sex, height, weight, and body mass index (BMI) (BMI, calculated as kg/m²), were recorded for all participants. Clinical information, including disease duration, current medications, and H-Y stage, was documented in medical records and during clinical examination.

Assessment Procedures

All assessments were performed while participants were in the "on" medication state to evaluate typical functional capacity. Measurements were performed by a single trained researcher using standardized protocols to ensure consistency and reliability. Given that all assessments were conducted by a single researcher following standardized protocols with prior training in each measure, intra-rater consistency was maintained throughout the study.

Disease severity was assessed using the H-Y staging scale, a widely used clinical tool that classifies PD progression based on motor symptom distribution and functional impact. The scale ranges from stage 1 (unilateral symptoms with minimal functional impairment) to stage 5 (wheelchair-bound or bedridden). This study included patients in stages 1-4, representing early to moderate-advanced disease (12).

The SARC-F was selected as the primary screening tool because it is a brief, validated instrument specifically recommended for case-finding in clinical settings by EWGSOP2, and because dual X-ray absorptiometry (DXA) or bioelectrical impedance analysis (BIA) muscle mass measurement was not feasible in this outpatient setting. (3,13). It consists of five items assessing strength, falls, the ability to rise from a chair, the need for assistance with walking, and the ability to climb stairs. Each element is scored

0-2 points, yielding a total score of 0-10. Scores ≥4 indicate high sarcopenia risk and demonstrate good sensitivity and specificity for identifying individuals requiring further assessment (13). Importantly, SARC-F identifies individuals at risk who require further comprehensive assessment, and results are herein interpreted as sarcopenia risk rather than a confirmed sarcopenia diagnosis (3,13). The Turkish version of the SARC-F has been validated for use in older adults.

Functional mobility was evaluated using the TUG test, a valid and reliable assessment tool for functional mobility widely used in PD populations (14). Participants were seated in a standard chair with armrests. Upon the instruction "go," participants rose from the chair, walked three meters at their comfortable pace, turned 180 degrees, returned to the chair, and sat down. Time to complete the task was recorded in seconds using a digital stopwatch. The test was demonstrated to participants prior to the actual measurement, and participants were allowed one practice trial. Timed Up and Go times >12-13 seconds are generally associated with increased fall risk and mobility impairment in PD populations (14,15).

Safety precautions included positioning a researcher near the participant throughout testing to provide assistance as needed. Participants were instructed to use their usual walking aids when appropriate. Adequate rest periods (at least 5 minutes) were provided between assessments to prevent fatigue.

Statistical Analysis

Statistical analyses were conducted using IBM SPSS Statistics version 25.0 (IBM Corp., Armonk, NY, USA). Normality was assessed using the Kolmogorov-Smirnov test. Descriptive statistics are reported as mean ± standard deviation for continuous variables and as frequency (percentage) for categorical variables.

Participants were categorized into two groups based on SARC-F scores: high sarcopenia risk (SARC-F ≥4) and low sarcopenia risk (SARC-F <4). Between-group comparisons for continuous variables were performed using an independent t-test for normally distributed variables (age, BMI, height, weight) and the Mann-Whitney U test for non-normally distributed variables (SARC-F score, TUG time). Categorical variables were compared using Fisher's exact test.

Relationships among SARC-F score, H-Y stage, and TUG time were assessed using Spearman's rank correlation coefficient because of the ordinal nature of H-Y staging and potential non-normal distributions. Ninety-five percent [95% confidence intervals (CIs)] were calculated for key correlation coefficients to enhance statistical reporting. Correlation strength was interpreted as follows: 0.00-0.29 (weak), 0.30-0.49 (moderate), 0.50-0.69 (strong), and ≥0.70 (very strong). Due to the relatively small sample size, multivariable regression analysis was not performed, and the analyses were limited to bivariate comparisons and correlation analyses.

Statistical significance was set at $p<0.05$ (two-tailed) for all analyses. All tests were conducted with appropriate consideration of effect sizes and clinical meaningfulness alongside statistical significance.

Results

Participant Characteristics

A total of 29 individuals with idiopathic PD (17 women, 12 men) participated in this study. The mean age was 70.59±12.61 years, and the mean BMI was 24.67±1.75 kg/m². Women comprised 58.6% of the sample. The mean SARC-F score was 4.00±1.93 points, ranging from 2 to 8. Using the SARC-F cut-off of ≥4, 19 participants (65.5%) were classified as high-risk for sarcopenia, and 10 (34.5%) as low. The distribution of H-Y stages was as follows: stage 1 (n=8, 27.6%), stage 2 (n=14, 48.3%), stage 3 (n=4, 13.8%), and stage 4 (n=3, 10.3%). The mean TUG time for the cohort was 24.38±11.54 seconds. The distribution of SARC-F and TUG scores across H-Y stages is illustrated in Figure 2.

Comparison Between Sarcopenia Risk Groups

Table 1 compares demographic and clinical characteristics of the high- and low-sarcopenia-risk groups. Participants in the high-risk group were significantly older than those in the low-risk group (75.42±10.36 vs. 63.00±11.26 years; p=0.008). The high-risk group also had significantly more advanced disease severity, with a mean H-Y stage of 2.37±0.83, compared with 1.50±0.53 in the low-risk group (p=0.005).

Functional mobility, as assessed by the TUG test, differed markedly between groups. The high-risk group took significantly longer to complete the TUG test (27.58±12.24 seconds) than the low-risk group (18.30±3.77 seconds; p=0.018). Notably, 73.7% (n=14) of participants in the high-risk group exceeded the 13-second TUG threshold, which is commonly considered indicative of increased fall risk in PD populations, whereas only 50.0% (n=5) of the low-risk group did so.

The gender distribution showed a trend toward a higher sarcopenia risk among women, with 70.6% (12/17) of women classified as high risk compared with 58.3% (7/12) of men; however, this difference was not statistically significant (p=0.694). No significant differences were observed in BMI between the two groups (24.70±1.80 vs. 24.61±1.71 kg/m²; p=0.902).

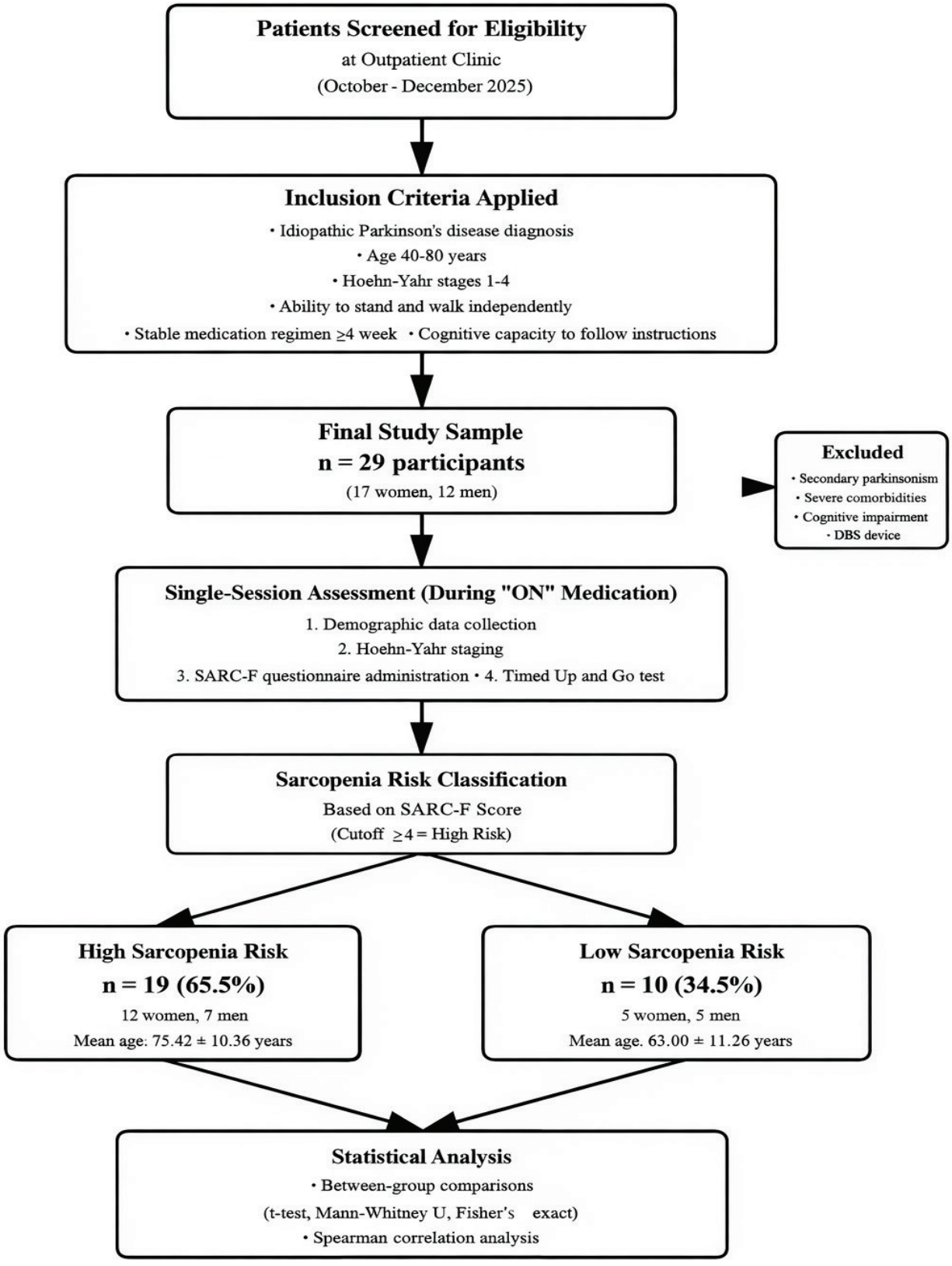

Figure 1. Study flowchart

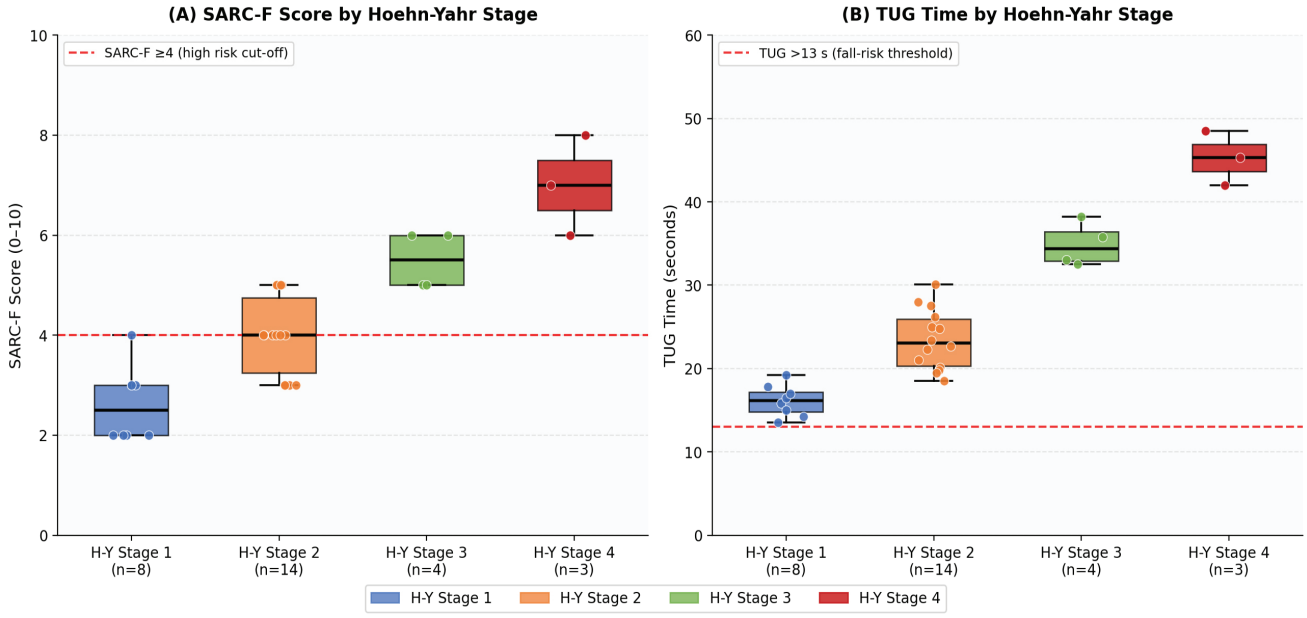

Figure 2. Box plots show median (horizontal line), interquartile range (box), and 1.5 × IQR (whiskers). Individual data points are overlaid. Red dashed lines indicate the SARC-F ≥4 high-risk cut-off (Panel A) and the TUG >13-second fall-risk threshold (Panel B)
SARC-F: Strength, assistance with walking, rise from a chair, climb stairs, and falls questionnaire, TUG: Timed Up and Go, H-Y: Hoehn-Yahr, IQR: Interquartile range

Table 1. Demographic and clinical characteristics of study participants

Variable	Total (n=29)	High-risk (n=19)	Low-risk (n=10)	p-value	Statistical test
Age (years)	70.59±12.61	75.42±10.36	63.00±11.26	0.008	Independent t-test
Gender, n (%)				0.694	Fisher's exact test
- Female	17 (58.6%)	12 (63.2%)	5 (50.0%)		
- Male	12 (41.4%)	7 (36.8%)	5 (50.0%)		
Height (cm)	164.24±6.54	163.47±7.12	165.70±5.31	0.384	Independent t-test
Weight (kg)	68.28±8.28	67.74±8.95	69.40±7.14	0.619	Independent t-test
BMI (kg/m²)	24.67±1.75	24.70±1.80	24.61±1.71	0.902	Independent t-test
SARC-F score	4.00±1.93	5.37±1.54	2.70±0.48	<0.001	Mann-Whitney U test
H-Y stage	1.97±0.85	2.37±0.83	1.50±0.53	0.005	Mann-Whitney U test
TUG time (seconds)	24.38±11.54	27.58±12.24	18.30±3.77	0.018	Mann-Whitney U test

Data presented as mean ± standard deviation or frequency (percentage)
BMI: Body mass index, H-Y stage: Hoehn-Yahr stage, TUG: Timed Up and Go

Correlation Analysis

Spearman correlation analysis revealed significant relationships among sarcopenia risk, disease severity, and functional mobility (Table 2). The SARC-F score showed a moderate positive correlation with H-Y stage (r=0.440, 95% CI: 0.08-0.70, p=0.016), indicating that a higher sarcopenia risk was associated with greater disease severity. A very strong positive correlation was observed between the SARC-F score and TUG time (r=0.709, 95% CI: 0.44-0.87, p<0.001). The relationship between disease severity and functional mobility was also robust, with H-Y

Table 2. Spearman correlation coefficients among study variables

Variables	r	p-value
SARC-F score × H-Y stage	0.440	0.016
SARC-F score × TUG time	0.709	<0.001
H-Y stage × TUG time	0.774	<0.001
Age × SARC-F score	0.427	0.020
Age × H-Y stage	0.377	0.044
Age × TUG time	0.477	0.009

All correlations calculated using Spearman's rank correlation coefficient
H-Y stage: Hoehn-Yahr stage, TUG: Timed Up and Go

stage showing a very strong positive correlation with TUG time (r=0.774, 95% CI: 0.55-0.90, p<0.001). Additionally, age showed a moderate positive correlation with the SARC-F score (r=0.427, p=0.020), confirming the age-related nature of sarcopenia risk in PD patients.

Discussion

The principal findings of this cross-sectional study demonstrate that 65.5% of PD patients are at high-risk of sarcopenia based on SARC-F screening; patients in this high-risk group are significantly older, have more advanced disease severity, and have substantially impaired functional mobility compared with the low-risk group. Very strong correlations were observed between sarcopenia risk and TUG performance (r=0.709) and between H-Y stage and TUG time (r=0.774), indicating that both disease progression and sarcopenia risk independently contribute to functional mobility impairment. To our knowledge, this is one of the first studies conducted in a Turkish PD cohort to demonstrate a strong quantitative relationship between SARC-F-based sarcopenia risk and objective functional mobility performance measured by the TUG test.

The observed sarcopenia risk prevalence of 65.5% in this cohort is notably higher than many previously reported estimates but falls within the broad range documented in the literature (5-7). Several factors likely contribute to this elevated estimate. First, the SARC-F questionnaire is a screening tool that identifies individuals at risk who require further assessment rather than confirming sarcopenia through comprehensive body-composition analysis. Therefore, findings should be interpreted as indicating sarcopenia risk rather than confirmed sarcopenia. Studies using SARC-F in PD populations have reported highly variable results: one study found 20% with confirmed sarcopenia and 34.5% screening positive on SARC-F (16), while others reported 47.2% probable sarcopenia and 55.5% screening positive using SARC-F (17,18). Second, the cohort's mean age (70.59 years), and particularly the high-risk group's mean age (75.42 years), indicate an elevated risk for both age- and disease-related muscle dysfunction. Age-stratified analyses reveal that sarcopenia prevalence in PD patients aged ≥70 years is substantially higher than in younger patients (8). Third, the SARC-F questionnaire, while demonstrating high specificity (typically 83-94%), exhibits relatively low sensitivity (13-31%) in many populations (13,16), yet paradoxically may capture functional limitations in PD that extend beyond pure sarcopenia to encompass disease-specific motor impairments such as bradykinesia and rigidity. This phenomenon, whereby PD-related motor impairment inflates SARC-F scores beyond what would be expected from muscle loss alone, represents a critical methodological consideration when interpreting SARC-F-based prevalence estimates in PD cohorts.

The high-risk group exhibited significantly more advanced H-Y staging (2.37 vs. 1.50, p=0.005), aligning with studies indicating that muscle strength diminishes and sarcopenia prevalence escalates with disease progression (8,9). Nevertheless, the correlation between sarcopenia and disease duration or severity is inconsistent among studies, with some indicating robust associations while others report no significant correlation (5). This variability probably shows how many different things can cause muscle problems in people with PD. In addition to disease-specific motor impairments such as bradykinesia, rigidity, and postural instability, other factors include decreased physical activity, nutritional deficiencies, chronic inflammation, oxidative stress, and possible effects of dopaminergic treatment (4,19). A recent 2025 prospective cohort study revealed that sarcopenia was independently correlated with functional degeneration in PD, even after controlling for age, sex, and comorbidities. This indicates that sarcopenia constitutes a unique and supplementary pathway to functional decline, extending beyond the progression of motor symptoms alone (9). The significant age difference between the risk groups in our study (75.42 vs. 63.00 years) underscores that age-related sarcopenia mechanisms may compound disease-related muscle changes, creating a synergistic effect that disproportionately affects older PD patients.

The very strong correlation between the SARC-F score and TUG performance (r=0.709, 95% CI: 0.44-0.87, p<0.001) is among the study's most clinically significant findings. The high-risk group took 51% longer to complete the TUG test (27.58 vs. 18.30 seconds), with mean times substantially exceeding established fall-risk thresholds (14,15,20). Previous research has confirmed that sarcopenia in PD is significantly associated with increased fall risk, with individuals screening positive on SARC-F experiencing falls more frequently than those screening negative (18,20,21). The current findings extend this evidence by demonstrating, in a Turkish PD cohort, that the magnitude of sarcopenia risk, as measured by SARC-F, is linearly related to TUG performance. This suggests that SARC-F is not merely a binary risk classifier but provides clinically meaningful graded information about functional mobility status in PD.

The robust correlation between H-Y stage and TUG time (r=0.774, 95% CI: 0.55-0.90, p<0.001) indicates that disease progression directly compromises functional mobility, consistent with established understanding of PD motor symptom evolution (12). The additional effect of sarcopenia risk on TUG performance—beyond disease severity alone—suggests that muscle dysfunction is an

independent contributor to mobility impairment. This has important clinical implications: while disease-modifying treatments for PD remain limited, sarcopenia is a potentially modifiable factor through targeted exercise interventions (10,11). Given that fall risk increases substantially with declining functional mobility, the combined use of SARC-F and TUG as a brief, non-instrumental screening battery could be integrated into routine neurological outpatient assessments without additional resources or time burden.

Screening should be integrated into routine neurological assessments, with positive results prompting referral to physical therapy or rehabilitation services. Exercise-based interventions, including progressive resistance training, balance exercises, and functional mobility training, have demonstrated efficacy in improving muscle strength and physical performance in PD populations (10,11,22,23). A recent Cochrane systematic review and network meta-analysis found that resistance and endurance training combined with balance exercises are some of the best ways to improve motor function in people with PD. This shows how important it is to find sarcopenia early on to help plan rehabilitation (10).

Study Limitations

Several limitations of this study must be acknowledged. First, the small sample size (n=29) and single-center design substantially limit the generalizability of findings to broader PD populations. Subgroup comparisons and correlation analyses performed with small sample sizes may be statistically fragile and should therefore be interpreted with caution. Future studies with larger samples are needed to confirm these findings. Second, sarcopenia risk was assessed using SARC-F screening without confirmatory muscle mass measurement by DXA or BIA, precluding a definitive diagnosis of sarcopenia according to EWGSOP2 criteria (3). Accordingly, results should be interpreted strictly as "sarcopenia risk" rather than as confirmed sarcopenia, and the term "sarcopenia risk" is used consistently throughout this manuscript. Third, the cross-sectional design precludes assessment of longitudinal relationships or causal inferences; the direction of the association between sarcopenia and functional decline cannot be established from this study alone. Fourth, data on potential confounding factors including nutritional status, physical activity levels, medication dosages, disease duration, and comorbidities were not systematically collected, and multivariable regression adjusting for these covariates was not performed; this represents an additional limitation. Fifth, the study was conducted at a single tertiary-care center in Türkiye, which may not be representative of the general PD population, including patients managed in primary- or secondary-care settings. Finally, SARC-F scores in PD may be inflated by disease-specific motor features (bradykinesia and rigidity), potentially overestimating true sarcopenia risk. Another limitation is that there is no a priori sample size or power calculation, which may limit the statistical robustness of subgroup comparisons and correlation analyses. Despite these limitations, this study provides novel evidence from a Turkish PD cohort that SARC-F-based sarcopenia screening is strongly correlated with functional mobility and disease severity. The use of two practical, validated, and non-instrumental tools (SARC-F and TUG) represents a clinically implementable approach that requires no specialized equipment and is therefore applicable to resource-limited outpatient settings. The robust correlations observed across multiple clinical parameters strengthen the internal validity of the findings within this sample.

These findings carry direct clinical implications. Routine incorporation of SARC-F screening and TUG assessment into PD outpatient follow-up visits requires minimal resources and can help identify patients at high-risk of functional decline who may benefit from timely referral to rehabilitation services. Given that sarcopenia is a modifiable condition, early detection could trigger preventive exercise interventions before significant functional loss occurs. Future research should use longitudinal designs with larger, multicenter samples and include objective diagnostic criteria for sarcopenia (DXA or BIA for muscle mass; handgrip strength for muscle strength) to clarify the temporal relationships between sarcopenia development and PD progression and to determine whether sarcopenia independently predicts fall incidence and hospitalization in this population.

Conclusion

In this Turkish PD cohort, sarcopenia risk assessed by SARC-F was highly prevalent and strongly associated with both disease severity and impaired functional mobility. These findings support the routine integration of SARC-F and TUG screening into PD clinical practice to enable early identification of at-risk patients and timely initiation of targeted, exercise-based rehabilitation. Sarcopenia is a changeable factor that can lead to functional decline, unlike the underlying neurodegenerative process. This means that rehabilitation strategies that focus on muscle dysfunction can help people with PD lower their risk of falling and maintain their quality of life. Longitudinal multicenter studies utilizing objective diagnostic criteria for sarcopenia are necessary to validate these associations and inform evidence-based clinical recommendations.

Ethics

Ethics Committee Approval: The University of Health Sciences Türkiye, Istanbul Kanuni Sultan Suleyman Training and Research Hospital Institutional Clinical Research Ethics

Committee examined and approved the study protocol (approval number: KAEK/2025.09.235, date: 25.09.2025).

Informed Consent: Written informed consent was obtained from participants after comprehensively outlining the study's goal, procedures, potential benefits, and risks.

Footnotes

Authorship Contributions

Surgical and Medical Practices: H.K.A., D.U.O., Z.K.G., Concept: H.K.A., D.U.O., Design: H.K.A., D.U.O., Data Collection or Processing: H.K.A., Z.K.G., Analysis or Interpretation: M.Z., Z.K.G., Literature Search: M.Z., D.U.O., Writing: M.Z.

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

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

References

1. Bloem BR, Okun MS, Klein C. Parkinson's disease. Lancet. 2021;397:2284-303.
2. GBD 2016 Parkinson's disease collaborators. Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Neurol. 2018;17:939-53.
3. Cruz-Jentoft AJ, Bahat G, Bauer J, et al. Sarcopenia: revised European consensus on definition and diagnosis. Age Ageing. 2019;48:16-31.
4. Chiba I, Lee S, Bae S, et al. Difference in sarcopenia characteristics associated with physical activity and disability incidences in older adults. J Cachexia Sarcopenia Muscle. 2021;12:1983-94.
5. Ponsoni A, Sardeli AV, Costa FP, Mourão LF. Prevalence of sarcopenia in Parkinson's disease: a systematic review and meta-analysis. Geriatr Nurs. 2023;49:44-9.
6. Hart A, Cordova-Rivera L, Barker F, Sayer AA, Granic A, Yarnall AJ. The prevalence of sarcopenia in Parkinson's disease and related disorders- a systematic review. Neurol Sci. 2023;44:4205-17.
7. Murakami K, Koh J, Ogami S, et al. Prevalence, impact, and screening methods of sarcopenia in Japanese patients with Parkinson's disease: a prospective cross-sectional study. Cureus. 2024;16:e65316.
8. Cai Y, Feng F, Wei Q, Jiang Z, Ou R, Shang H. Sarcopenia in patients with Parkinson's disease: a systematic review and meta-analysis. Front Neurol. 2021;12:598035.
9. Liu G, Hong C, Xu S, et al. Association of sarcopenia with Parkinson's disease and related functional degeneration among older adults: a prospective cohort study in Europe. J Affect Disord. 2025;374:553-62.
10. Fan X, Yuan Y, Bai Y, et al. Optimal dose and type of exercise improve the overall balance in adults with Parkinson's disease: a systematic review and Bayesian network meta-analysis. Neurol Sci. 2025;46:4169-80.
11. Wang J, Yuan Y, Xie T, et al. Optimal dose and type of exercise to improve motor symptoms in adults with Parkinson's disease: a network meta-analysis. J Sci Med Sport. 2025;28:282-91.
12. Opara J, Małecki A, Małecka E, Socha T. Motor assessment in Parkinson's disease. Ann Agric Environ Med. 2017;24:411-5.
13. Malmstrom TK, Miller DK, Simonsick EM, Ferrucci L, Morley JE. SARC-F: a symptom score to predict persons with sarcopenia at risk for poor functional outcomes. J Cachexia Sarcopenia Muscle. 2016;7:28-36.
14. Nocera JR, Stegemöller EL, Malaty IA, Okun MS, Marsiske M, Hass CJ; National Parkinson Foundation Quality Improvement Initiative Investigators. Using the Timed Up & Go test in a clinical setting to predict falling in Parkinson's disease. Arch Phys Med Rehabil. 2013;94:1300-5.
15. Morris S, Morris ME, Iansek R. Reliability of measurements obtained with the Timed "Up & Go" test in people with Parkinson disease. Phys Ther. 2001;81:810-8.
16. Choi S, Kim R, Lee JY, Jeon B. Clinical usefulness of the SARC-F questionnaire to assess sarcopenia in patients with Parkinson's disease. Parkinsonism Relat Disord. 2024;125:107039.
17. Lima DP, de Almeida SB, Bonfadini JC, et al. Clinical correlates of sarcopenia and falls in Parkinson's disease. PLoS One. 2020;15:e0227238.
18. Peball M, Mahlknecht P, Werkmann M, et al. Prevalence and associated factors of sarcopenia and frailty in Parkinson's disease: a cross-sectional study. Gerontology. 2019;65:216-28.
19. Barichella M, Cereda E, Ferri V, et al. Sarcopenia, low muscle strength, cognitive functions, and quality of life in parkinsonian syndromes. Nutrition. 2024;128:112568.
20. Mollà-Casanova S, Pedrero-Sánchez J, Inglés M, et al. Impact of Parkinson's disease on functional mobility at different stages. Front Aging Neurosci. 2022;14:935841.
21. Rajalingam R, Sorrento G, Fasano A. Risk of fall with device-based advanced treatments in Parkinson's disease: a systematic review and network meta-analysis. J Neurol Neurosurg Psychiatry. 2025;96:470-9.
22. Goldman JG, Volpe D, Ellis TD, et al. Delivering multidisciplinary rehabilitation care in Parkinson's disease: an international consensus statement. J Parkinsons Dis. 2024;14:135-66.
23. Zeng B, Wang H, Chen J, Wu S, Xie J. Optimal exercise modalities and doses for improving quality of life in patients with Parkinson's disease: a network meta-analysis and dose-response study. J Neurol. 2026;273:123.