e-ISSN: 2149-3189

European Research Journal

Volume 11 Issue 5 September 2025

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

Published by Prusa Medical Publishing

The European Research Journal

Aim and Scope

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

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

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

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

The journal is published bimonthly (January, March, May, July, September, and November).

Abstracting and Indexing

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

Publisher

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

https://dergipark.org.tr/en/pub/eurj
https://www.prusamp.com

e-ISSN: 2149-3189

EDITORIAL BOARD

EDITOR-IN-CHIEF

EDITORS

OWNER ON BEHALF OF THE PRUSA MEDİCAL PUBLISHING

ASSISTANT EDITOR

SECTION EDITORS

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

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

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

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

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

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

Nurullah DOGAN, MD.,
Associate Professor,
Doruk Nilüfer Hospital,
Department of Radiology,
Bursa, Turkey

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

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

LANGUAGE EDITOR

ETHICAL EDITOR

INTERNATIONAL EDITORIAL BOARD MEMBERS

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

Yenal DUNDAR, MD.,
Consultant Psychiatrist
Central Queensland Hospital and Health Service,
QLD, Australia

Table of Contents

Original Articles

The European Research Journal 2025;11(5):832-844 Original Article

DOI: https://doi.org/10.18621/eurj.1700331 *Anatomy*

Morphometric analysis of anatomical reference points in hip surgery: A study on cadaveric and radiographic images

Ayşe Gül Kabakcı[1], Ömer Sunkar Biçer[2], Memduha Gülhal Bozkır[1]

[1]Department of Anatomy, Cukurova University, Faculty of Medicine, Adana, Türkiye; [2]Department of Orthopaedics and Traumatology, Cukurova University, Faculty of Medicine, Adana, Türkiye

ABSTRACT

Objectives: This study aims to evaluate the accuracy of morphometric measurements obtained from cadavers and anteroposterior (AP) pelvic radiographs.

Methods: A total of 15 cadavers from the anatomical collection of Cukurova University and 217 AP pelvic radiographs from individuals aged 65-80 years with no orthopedic conditions were analyzed. Morphometric measurements were taken from cadavers using Kirschner wires placed at anatomical reference points: the spina iliaca anterior superior (SIAS), the highest point of the crista iliaca (CI), and the trochanter major (TM). Distances were measured with a non-elastic tape, and angular measurements were conducted using ImageJ software. These were compared with radiographic data analyzed via the PACS system. Statistical analysis was performed using SPSS 20 with One-Way ANOVA to assess group differences.

Results: The mean age was 70.98±4.70 years for radiograph individuals and 80.36±5.13 years for cadavers. Significant differences were found between cadaveric, dissected cadaver, and radiological measurements. The SIAS-TM distance was longest in cadavers (113.82±7.46 mm) and shortest in radiographs (92.73±14.36 mm). The CI-TM distance was greatest in radiographs (147.81±12.02 mm), while the SIAS-CI distance was longest in cadavers (78.95±6.48 mm). Differences in SIAS-TM ($P<0.001$), CI-TM ($P=0.007$), and SIAS-CI ($P=0.029$) distances were statistically significant. Angular measurements also varied, with radiographs showing higher SIAS angles and cadavers showing greater TM and CI angles, especially on the right side.

Conclusions: The study reveals notable discrepancies between cadaveric and radiological morphometric measurements. These cadaver-based findings may serve as valuable resources for surgical training and anatomical education, especially in hip arthroplasty planning.

Keywords: Arthroplasty, cadaver, hip, morphometry, radiography

Total hip arthroplasty (THA) is a significant surgical procedure performed to improve joint stability, mobility, gait function, and overall quality of life. In orthopedic practice, femoral neck fractures and hip prosthesis applications are commonly addressed using various techniques, including the anterolateral approach. In recent years, the use of minimally invasive techniques in THA has led to no-

Received: May 15 2025 **Accepted:** June 27, 2025 **Available Online:** July 15, 2025 **Published:** September 4, 2025

How to cite this article: Kabakcı AG, Biçer ÖS, Bozkır MG. Morphometric analysis of anatomical reference points in hip surgery: A study on cadaveric and radiographic images. Eur Res J. 2025;11(5):832-844. doi: 10.18621/eurj.1700331

Corresponding author: Ayşe Gül Kabakçı, PhD., Phone: +90 322 338 60 60 ext. 3489, E-mail: akabakci@cu.edu.tr

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

table improvements in surgical outcomes. Nevertheless, standard surgical approaches remain widely used and continue to yield generally favorable results [1]. The preference for minimally invasive techniques accelerates the recovery process and enhances soft tissue healing in various surgical procedures such as cholecystectomy, meniscectomy, and anterior cruciate ligament reconstruction. These positive developments have also made an impact in hip surgery and are increasingly being adopted in total hip arthroplasty procedures. Studies in the literature emphasize that minimally invasive techniques are associated with fewer complications and faster recovery processes [2-4]. However, interindividual anatomical variations and the technical diversity of orthopedic surgery make it important to accurately determine the distances between anatomical reference points during surgical planning. Aligning incision lines determined through clinical experience with precise anatomical measurements may enhance both the accuracy and safety of surgical procedures. Moreover, well-positioned and minimally sized incisions can accelerate recovery, reduce postoperative pain, and help patients regain functional mobility and return to daily life more quickly. Accordingly, the study sought to answer the following key questions: What is the degree of similarity between morphometric measurements obtained using Kirschner wires on cadavers and anthropometric data derived from anteroposterior (AP) pelvic radiographs? To what extent do the distances between anatomical reference points identified on radiological images correspond with the incision lines subjectively determined by orthopedic surgeons on cadavers? Furthermore, can morphometric measurements performed on cadavers, when compared with AP pelvic radiographic data, lead to more precise and safer approaches in surgical practice? Based on these questions, the study was designed around the following hypotheses:

(1) There is a significant difference between Kirschner wire measurements performed on cadavers and anthropometric measurements obtained from AP pelvic radiographs.

(2) Kirschner wire measurements made on cadavers are similar to measurements obtained after dissection.

Morphometric analyses in the literature are mostly based on computed tomography (CT) images [5, 6] and dry bone models [7, 8]. However, these methods have limitations in reflecting soft tissue relationships, which may reduce their clinical applicability. Furthermore, many cadaver-based morphometric studies in the literature primarily focus on the proximal femur, particularly the femoral head, neck, and trochanteric region which are directly involved in prosthesis placement [9, 10]. However, effective surgical incision planning also requires precise knowledge of the upper anatomical boundaries of the pelvis, including palpable and clinically relevant landmarks such as the spina iliaca anterior superior (SIAS), the highest point of the crista iliaca (CI), and the trochanter major (TM).

Despite their importance, cadaveric measurements involving these pelvic landmarks remain limited in the current literature. This study aims to address this gap by providing a comparative analysis of morphometric data obtained from both cadavers and anteroposterior pelvic radiographs, focusing specifically on these anatomical reference points. In doing so, the study not only contributes to the academic body of knowledge but also supports the development of safer and more reliable incision strategies in surgical practice. Moreover, another aim of this study is to contribute to surgical training programs focused on hip arthroplasty by providing detailed cadaveric measurements and anatomical images of the pelvic region. Cadaver-based training plays a critical role in enhancing surgeons' confidence and offering a safe environment for practicing complex surgical procedures [11-13]. These materials, which reflect real tissue characteristics, help reinforce anatomical knowledge and improve technical skills, thereby facilitating a more effective transition from theoretical education to clinical practice.

In this context, our study aimed to evaluate morphometric measurements performed on cadavers in conjunction with radiological image analysis, to contribute to the development of anatomical standards that can be applied in clinical practice. Moreover, by providing measurable data that can support minimally invasive hip prosthesis approaches, this research lays a foundation for the refinement of surgical techniques.

METHODS

Study Population

In this study, 15 cadavers from the anatomical collection of the Department of Anatomy, Faculty of Medi-

cine, Cukurova University, and 217 anteroposterior (AP) pelvic radiographs were analyzed. Both the cadavers and the radiographic images belonged to individuals aged between 65 and 80 years. To minimize the confounding effect of age, we specifically selected radiographic images from individuals in the same age range as the cadavers (average age of cadavers: 80.36±5.13 years; radiograph group: 70.98±4.70 years). This approach ensured age homogeneity across groups, thereby eliminating the need for stratified analysis by age decades. This methodological decision was made to focus the analysis on anatomical and measurement technique differences, rather than demographic variability. The selected individuals had no history of orthopedic diagnosis or surgical intervention involving the lower extremities. Radiological data were obtained from the archives of the Department of Orthopedics and Traumatology at Faculty of Medicine, Cukurova University, following the necessary institutional approvals. Ethical approval for the use of cadavers was granted by the relevant institutional authorities, and the entire study was approved by the Non-Invasive Clinical Research Ethics Committee of Cukurova University Faculty of Medicine on November 8, 2024 (Decision No: 149/7). All experimental procedures were conducted in accordance with the principles of the Declaration of Helsinki.

Study Design

This study aims to determine the morphometric values of anatomical reference points important for hip arthroplasty. To enhance the clarity and reproducibility of the methodology, a structured summary of the step-by-step measurement protocol is presented in Table 1. This protocol encompasses both cadaveric and radiographic procedures, including details on subject preparation, anatomical reference point identification, measurement techniques, and the comparison of collected data. Cadaveric measurements involving the use of Kirschner wires were performed by a professor of

Table 1. Summary of morphometric measurement protocol

Step	Procedure	Details
Study objective	Determination of anatomical distances and comparison of measurement methods	Focus on SIAS, CI, and TM points relevant for hip arthroplasty
Cadaver preparation and marking	Identification and marking of reference points on cadavers (Figs. 1 and 2)	Kirschner wires placed by an orthopedic surgeon at SIAS, CI, and TM according to the hip incision line
Cadaveric measurements	Measurement of distances on cadavers (Figs. 1 and 2)	Non-elastic tape used to measure: - SIAS-TM - CI-TM - SIAS-CI Performed before and after dissection
Radiographic image acquisition	Retrieval and standardization of AP pelvic radiographs	- Supine patient positioning - Neutral lower limb alignment - 100 cm source-to-image distance (SID) - Radiopaque calibration marker placed at TM level
Radiographic measurements	Digital measurement on radiographs	DICOM images analyzed using PACS system tools; same anatomical points measured digitally
Data comparison	Cross-method analysis	Comparison of cadaveric and radiographic data to assess correlation, accuracy, and reproducibility

SIAS=spina iliaca anterior superior, TM=trochanter majör, CI=crista iliaca, PACS=Picture Archiving and Communication System, DICOM=Digital Imaging and Communication in Medicine

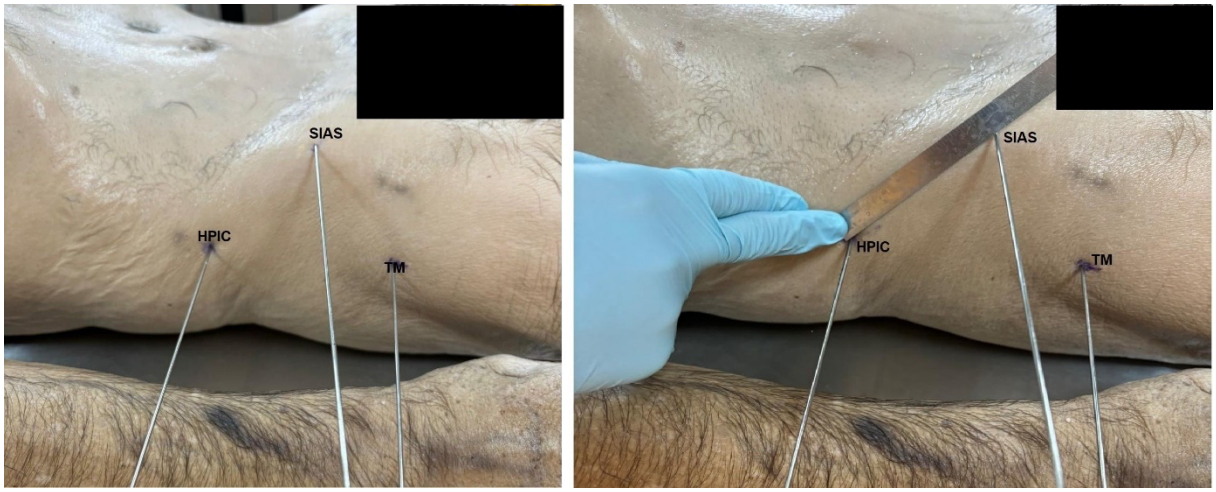

Fig. 1. Representation of the reference points (spina iliaca anaterior superior (SIAS), trochanter major (TM), and crista iliaca (CI)) and length measurements on cadaveric specimens.

orthopedics. The dissections of the cadavers were conducted by an experienced anatomist, who also carried out the post-dissection measurements as well as all radiographic measurements. As each type of measurement was performed by a single expert observer, interobserver reliability (e.g., Intraclass Correlation Coefficients) was not calculated. While this may limit the assessment of reproducibility across multiple raters, it ensures internal consistency by avoiding interobserver variability. Furthermore, the advanced qualifications and relevant expertise of both observers support the methodological reliability of the measurements.

Anatomical Reference Points and Measurements

In this study, the SIAS, the highest point of the CI, and the TM were identified as anatomical reference points. Length and angular measurements were conducted between these landmarks.

Cadaveric Morphometric Measurements

The distances between anatomical reference points were determined on cadavers using Kirschner wires placed by an orthopedic specialist. The wires were inserted by palpating the relevant anatomical landmarks - SIAS, the highest point of CI, and TM - with consideration of the planned incision line (Figs. 1 and 2). Following the placement of the wires, length measurements were performed between the reference points. In addition to these initial assessments, the same morphometric parameters were re-measured on the same cadavers after dissection, allowing for comparison between pre- and post-dissection values.

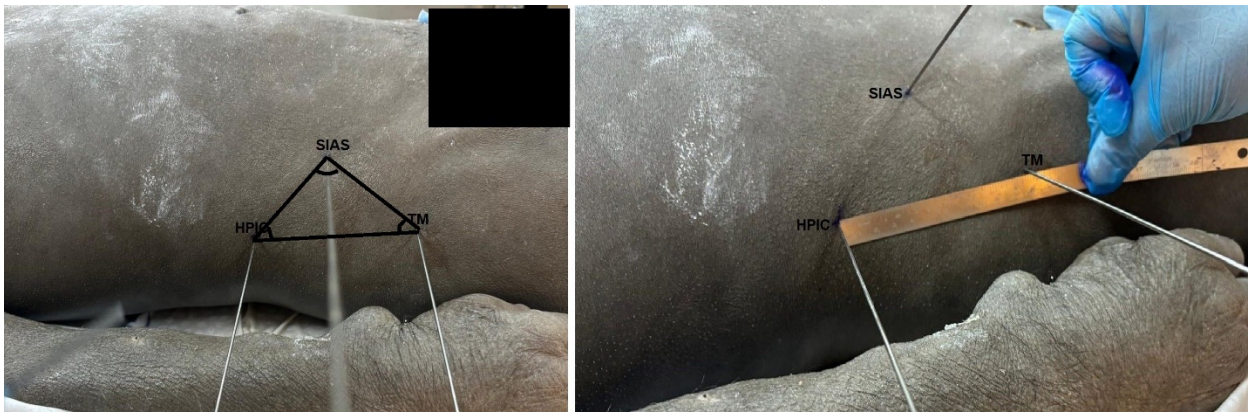

Fig. 2. Representation of the reference points (spina iliaca anterior superior [SIAS], trochanter major [TM], and crista iliaca [CI]) and angle measurements on cadaveric specimens.

Length Measurements

The distances between SIAS and TM, CI and TM, and SIAS and CI were measured using a non-elastic measuring tape to ensure consistency and accuracy (Fig. 1).

Angular Measurements

After transferring cadaveric images to a computer environment, angular measurements were digitally performed using ImageJ software (version 1.52a), which offers a precision of 1/100 mm. Using the angle measurement tool within the software, the distal point of the angle was first selected. A line was then drawn toward the vertex of the angle, followed by a second line extending to the third point, thus completing the angular configuration. The resulting angle values were recorded in degrees. This digital method enabled highly precise and repeatable angular assessments. Angular measurements were taken at the SIAS, CI, and TM points, as illustrated in Fig. 2.

Radiological Imaging and Morphometric Measurements

The same anatomical reference points - SIAS, the highest point of CI, and TM - were evaluated on anteroposterior (AP) pelvic radiographs using anthropometric measurements (Fig. 3). Radiographic images were acquired with patients in a supine position, with the lower limbs positioned neutrally and parallel to avoid rotation artifacts. A standard source-to-image distance (SID) of 100 cm was used during image acquisition, consistent with conventional pelvic radiography protocols. To ensure measurement accuracy, a radiopaque calibration marker (metallic ball of known diameter) was included at the level of the greater trochanter in each image field. This reference object allowed proper scaling and adjustment of digital measurements. Morphometric distances between SIAS and TM, CI and TM, and SIAS and CI were digitally measured using radiographic imaging software. Patient radiographs were retrieved from the Mergentech

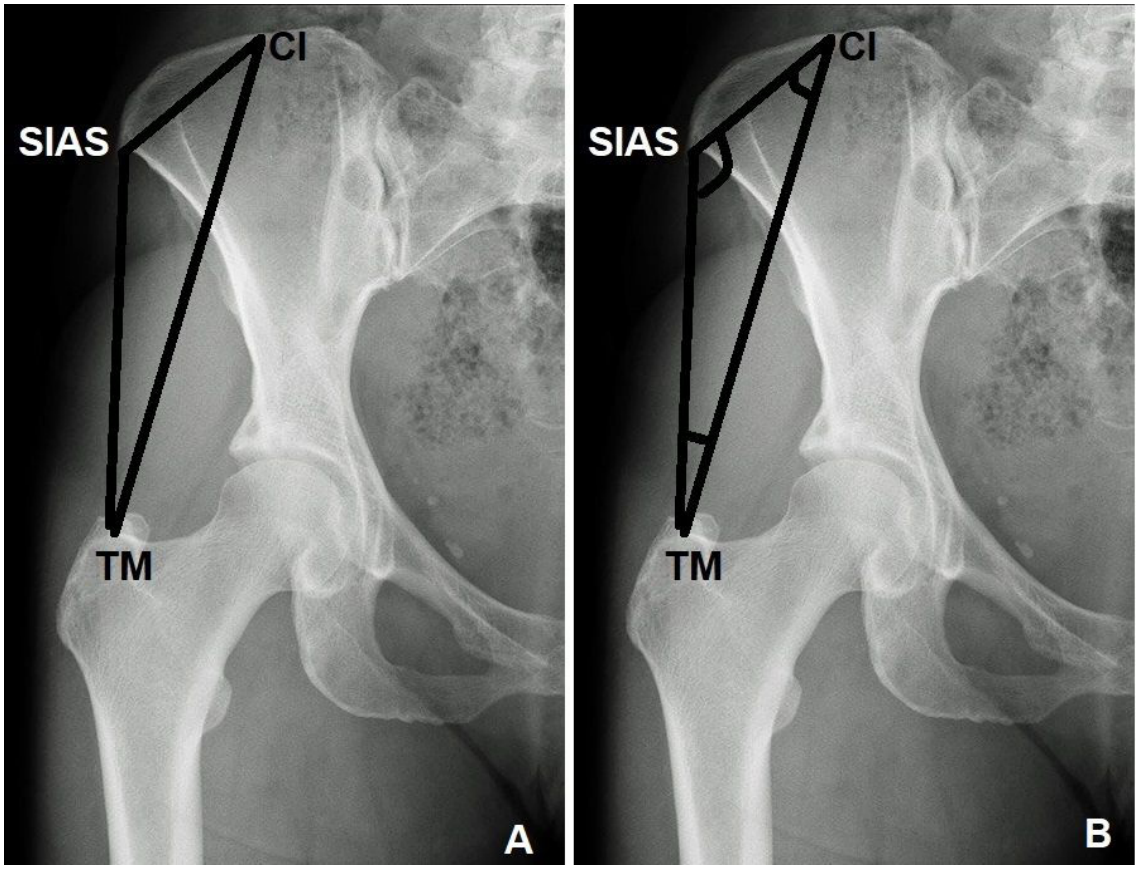

Fig. 3. Radiological representation of the reference points (spina iliaca anaterior superior [SIAS], trochanter major [TM], and crista iliaca [CI]) along with length and angle measurements

Hospital Information Management System (HIMS) and analyzed via the PACS (Picture Archiving and Communication System). All measurements were performed directly within the PACS environment on DICOM (Digital Imaging and Communication in Medicine)-format images, using its built-in measurement tools. PACS ensured high reliability and reproducibility through secure data handling, image fidelity, and measurement standardization. Morphometric data obtained from Kirschner wire placements on cadavers were compared with the corresponding radiological measurements to assess correlation and structural consistency between both methods. The imaging protocol and digital analysis tools used in this study are consistent with current best practices reported in radiologic morphometry literature.

Statistical Analysis

In our study, statistical analysis was performed on the morphometric data obtained from cadaveric and radiological images. Using SPSS version 20.0, the Descriptive Statistics function was utilized to calculate the mean and standard deviation of the morphometric parameters.

Prior to the application of parametric tests, the distribution of the data was evaluated using the Shapiro-Wilk test. The results indicated that the data were normally distributed across all measurement variables (e.g., SIAS–TM, CI–TM, SIAS–CI). For example, the Shapiro–Wilk test yielded the following P-values: (1) SIAS-TM (cadaveric) (P=0.421); (2) CI-TM (cadaveric) (P=0.386); (3) SIAS-CI (cadaveric) (P=0.537); (4) SIAS-TM (radiographic) (P=0.452); (5) CI-TM (radiographic) (P=0.496); and (6) SIAS-CI (radiographic) (P=0.478). As all P-values were greater than 0.05, the assumption of normality was met. Therefore, a One-Way Analysis of Variance (ANOVA) was conducted to compare the morphometric values between cadaveric and radiological measurements. This analysis revealed whether the differences between the two imaging modalities were statistically significant. Given the normal distribution of the data and fulfillment of test assumptions, the use of ANOVA was considered statistically valid and reliable.

RESULTS

A total of 15 cadavers (30 lower extremities; 11 males, 3 females) and 217 AP pelvic radiographs (434 lower extremities; 97 males, 120 females) were evaluated. The mean age of radiographic cases was 70.98±4.70 years, while cadavers averaged 80.36±5.13 years.

Morphometric comparisons revealed statistically significant differences across cadaveric, dissected cadaveric, and radiographic groups (Table 2) (Fig. 4). Notably, the SIAS-TM distance showed a progressive decrease from cadavers to radiographic images (F=31.66, P<0.001), with cadavers exhibiting the highest values and radiographs the lowest. Conversely, the CI-TM distance was longest in radiographs (F=5.40, P=0.007). The SIAS-CI distance was shortest

Table 2. The comparison of measurements taken with the Kischner wire on cadavers, measurements taken after dissection, and measurements made on AP pelvic images according to side

Parameters (mm)	Cadavers (n=15)	Total (n=30)	Dissected cadavers (n=15)	Total (n=30)	AP pelvic images (n=217)	Total (n=434)
SIAS-TM (R)	114.00±5.51	113.82±7.46	104.20±4.92	104.17±6.59	93.79±14.76	92.73±14.36
SIAS-TM (L)	113.63±9.29		104.13±8.11		91.59±13.88	
CI-TM (R)	137.64±6.27	136.05±7.16	138.80±6.06	137.43±6.84	147.04±12.47	147.81±12.02
CI-TM (L)	134.46±7.92		136.07±7.51		148.62±11.51	
SIAS-CI (R)	78.00±8.12	78.95±6.48	79.80±7.91	80.13±6.49	70.05±11.28	71.93±12.22
SIAS-CI (L)	79.91±4.50		80.53±4.84		73.91±12.86	

Data are shown as mean±standard deviation. SIAS=spina iliaca anterior superior, TM=trochanter major, CI=crista iliaca, R=right, L=left, mm=millimeter, AP= anteroposterior

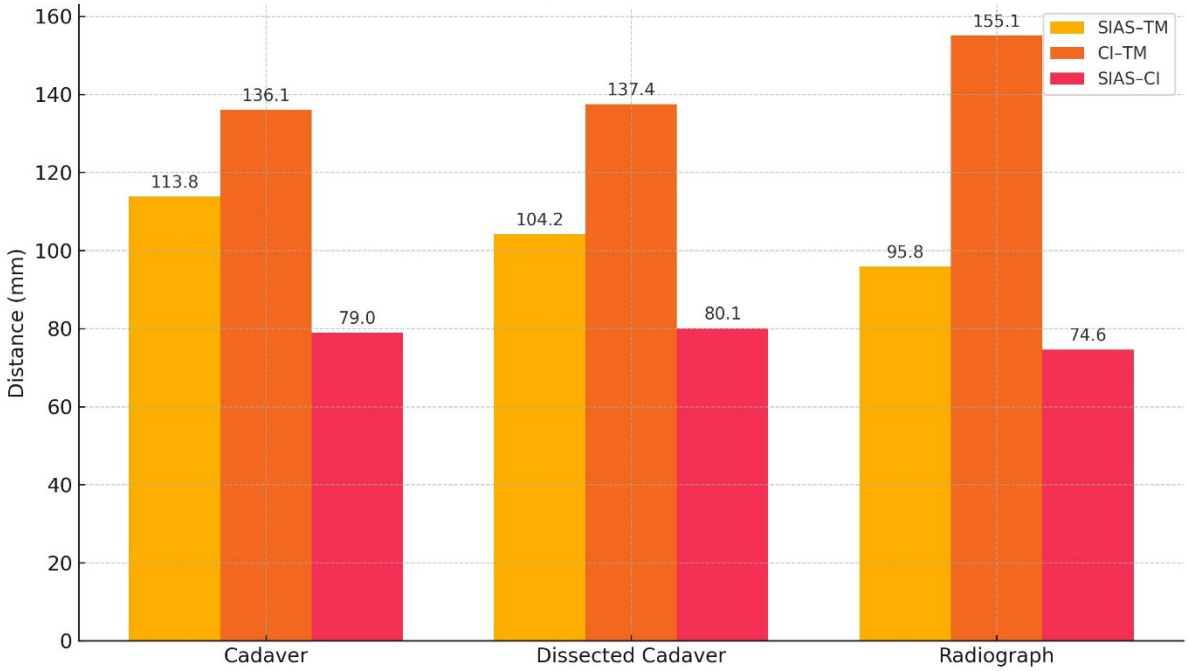

Fig. 4. Comparison of morphometric distances by metod

in radiographs and differed significantly between groups (F=3.76, P=0.029). These findings suggest that radiographic imaging may systematically alter linear measurements due to soft tissue compression or projection variability.

Laterality analysis revealed consistent patterns across all three groups. In cadaveric and dissected cadaveric specimens, the right-side measurements (e.g., SIAS-TM and CI-TM) were generally longer than the left, whereas in radiographic images, the difference between right and left was less pronounced and did not reach statistical significance. Interestingly, the SIAS-CI distance was consistently longer on the left side in all three groups, and this asymmetry was statistically significant. Moreover, angular evaluations at SIAS, TM, and CI landmarks also exhibited methodological variation. Radiographs yielded higher angular values at the SIAS, while TM and CI angles tended to be greater in cadaveric measurements, especially on the right side (Table 3, 4) (Fig. 5). These angular differences underline the influence of measurement modality on anatomical interpretation and may have clinical implications for surgical planning. When comparing the two methods, it was observed that radiological measurements tended to yield slightly higher angular values at the SIAS, whereas cadaveric measurements presented greater values at the TM and CI points, particularly on the right side. Moreover, Fig. 6

Table 3. Similarity between measurements marked with kischner wire, measurements taken post-dissection, and radiographic measurements

Parameters (mm)	Cadavers (n=30)	Dissected cadavers (n=30)	AP pelvic images (n=434)	F	P value
SIAS-TM	113.82±7.46	104.16±6.58	95.83±14.66	31.66	**<0.001**
CI-TM	136.05±7.16	137.43±6.84	155.13±13.89	5.40	**0.007**
SIAS-CI	78.95±6.48	80.13±6.48	74.58±12.34	3.76	**0.029**

Data are shown as mean±standard deviation. SIAS=spina iliaca anterior superior, TM=trochanter major, CI=crista iliaca, mm=millimeter, AP= anteroposterior

F=between-group variance; P=significance level (One-Way ANOVA)

Table 4. The angles at the measurement locations in cadavers and AP pelvic images

Parameters (mm)	Cadavers (n=15)	Total (n=30)	AP pelvic images (n=217)	Total (n=434)
SIAS (R)	121.27±9.13	119.04±7.46	130.38±4.58	126.53±6.20
SIAS (L)	116.43±4.28		120.77±1.07	
CI (R)	31.44±5.46	32.51±4.53	27.88±6.22	30.42±5.61
CI (L)	33.75±3.14		34.24±0.08	
TM (R)	28.77±4.70	28.47±3.88	24.70±2.03	25.80±2.58
TM (L)	28.12±3.06		27.45±3.04	

Data are shown as mean±standard deviation. SIAS=spina iliaca anterior superior, TM=trochanter major, CI=crista iliaca, R=right, L=left, mm=millimeter, AP= anteroposterior

presents 3D scatter plots illustrating the relationship among three key anatomical distances (CI-TM, SIAS-CI, and SIAS-TM) in radiological (A) and cadaveric (B) datasets. In the radiological group (A), data points are densely clustered and form a relatively linear distribution, indicating strong internal consistency and positive correlation among the three measured distances across a larger sample size. In contrast, the cadaveric group (B) displays a more dispersed and less densely packed pattern, reflecting greater interindividual variability likely due to anatomical diversity, postmortem changes, and lower sample size. Notably, while overall trends are preserved in both datasets, the cadaveric values occupy a relatively narrower CI-TM range and a higher SIAS-CI span, suggesting modality-specific influences on spatial relationships between anatomical landmarks.

DISCUSSION

The primary aim of this study is to contribute to the establishment of surgical standards by evaluating key anatomical reference points relevant to the incision line used in the minimally invasive anterolateral approach, through morphometric measurements per-

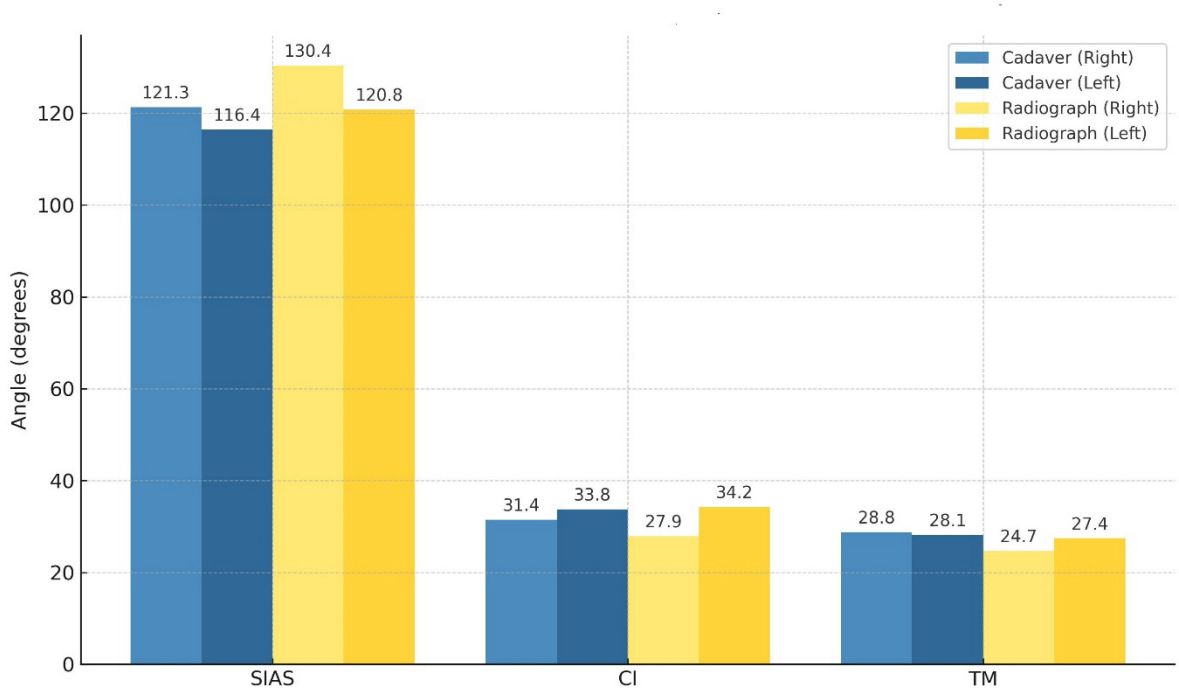

Fig. 5. Comparison of angular measurements by metod and side

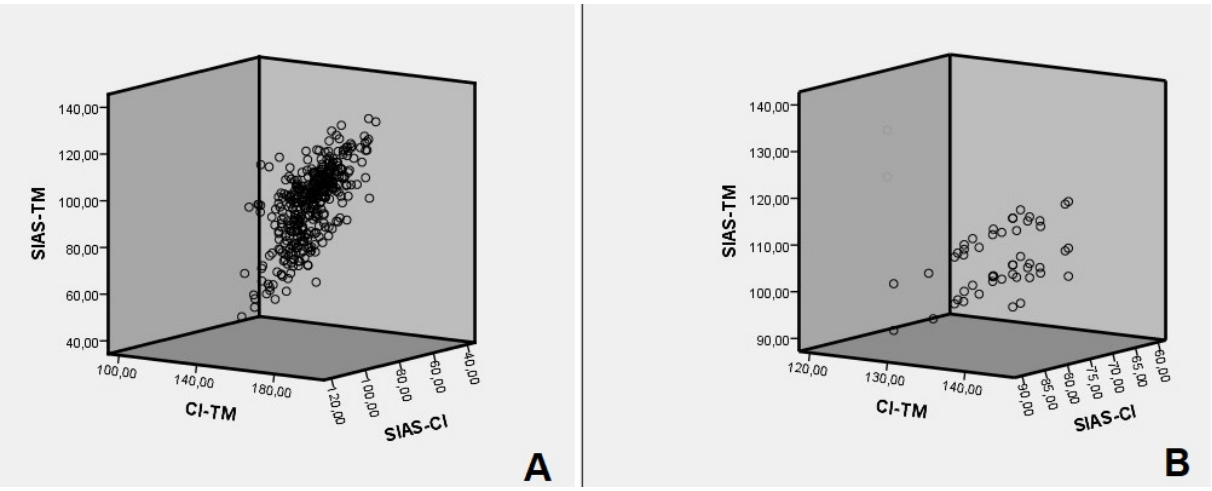

Fig. 6. Distribution of length measurements compared between radiological measurements (A) and cadaveric specimens (B)

formed on both cadavers and radiological images. In the literature, studies describing the minimally invasive anterolateral approach emphasize that the patient is positioned in the lateral decubitus position, and SIAS and TM are identified as critical anatomical landmarks. The incision is typically made approximately 10 cm in length, extending between the middle and posterior thirds of the line connecting these two points. In standard THA, the incision length typically ranges between 15 and 25 cm, whereas in minimally invasive techniques, this length is reduced to approximately 10 cm. This reduction in incision size highlights the importance of accurately identifying anatomical landmarks such as SIAS and TM to ensure both surgical efficacy and safety in limited surgical fields. Accordingly, our study focused on morphometric evaluation of the distances between these reference points on cadavers and anteroposterior pelvic radiographs, with the aim of providing objective data that can serve as reliable references in surgical practice.

The first hypothesis of this study stated that "there is a significant difference between Kirschner wire measurements performed on cadavers and anthropometric measurements obtained from AP pelvic radiographs." The findings of the study clearly support this hypothesis. Statistical analyses revealed significant differences among the cadaveric, dissected cadaver, and radiological measurements for the distances between SIAS-TM, CI-TM, and SIAS-CI ($P<0.05$; see Table 2 and Table 3). Notably, the SIAS-TM and SIAS-CI distances were consistently shorter in radiological measurements, which may be attributed to factors such as soft tissue compression, differences in positioning, and the limitations of two-dimensional imaging. These results confirm that the measurement method has a direct impact on the observed morphometric values, thereby validating the first hypothesis. The second hypothesis proposed that "Kirschner wire measurements made on cadavers are similar to measurements obtained after dissection." The results of this study also support this assumption. Comparisons between cadaveric and dissected cadaver data revealed no statistically significant differences for most of the morphometric parameters, and overall measurement trends were consistent. For instance, the CI-TM and SIAS-CI distances remained nearly identical before and after dissection, indicating that the dissection process did not substantially alter the spatial relationships of key anatomical landmarks. This finding is consistent with previous studies by Verma *et al*. [7] and Solomon *et al*. [8], which suggested that radiological images may underestimate anatomical distances, especially in complex regions like the hip. Such discrepancies highlight the potential limitations of relying solely on radiographic data for surgical planning, particularly in minimally invasive procedures. Therefore, cadaveric measurements, being more directly reflective of anatomical reality, are essential for ensuring precise surgical planning. On the other hand, other studies, such as those by Austin *et al*. [14], Vanrusselt *et al*. [15], Cantrell *et al*. [16], and Heinz *et al*. [17], have emphasized the importance of radiological eval-

uations, particularly in postoperative assessments. They argue that radiological morphometric measurements are crucial tools in evaluating the success of hip arthroplasty, providing vital insights into the positioning and alignment of the prosthesis post-surgery. Given these perspectives, it is clear that both cadaveric and radiological measurements offer valuable contributions to different stages of surgical planning and evaluation. Therefore, in our study, we integrated both types of measurements to ensure a comprehensive approach. By combining the accuracy of cadaveric measurements with the clinical relevance of radiological assessments, we aimed to provide a balanced and reliable framework for surgical planning, ensuring both precision in preoperative measurements and practical applicability in postoperative evaluations. This dual approach enhances the robustness of our findings and ensures that both anatomical accuracy and clinical feasibility are accounted for in the surgical process. Additionally, the study found differences in the side-to-side measurements, particularly in the SIAS-TM distance, where the right side was consistently longer in cadavers. In contrast, no significant side-to-side difference was found in radiological images. This mirrors findings from Sugano *et al*. [5], and Zhai *et al*. [6], who highlighted that anatomical asymmetries are common but often underrepresented in radiological imaging. This suggests that when planning surgeries, particularly in minimally invasive approaches, surgeons should account for potential anatomical variations between the sides, which may not be captured accurately in radiographs. Another noteworthy observation is the similarity between the measurements obtained from cadavers and dissected cadavers. The close alignment of these two measurement methods emphasizes the reliability of cadaveric measurements in surgical applications. As noted by Verma *et al*. [7], dissected cadavers provide an invaluable source of precise anatomical data that can be directly applied to surgical planning. This study further confirms that dissected cadavers are useful models for obtaining measurements with a high degree of accuracy, supporting their use in surgical training and preoperative planning.

Our study also revealed that the distances between anatomical points, such as the SIAS-TM and SIAS-CI, differed significantly between cadavers, dissected cadavers, and radiological images. Specifically, while the SIAS-TM and SIAS-CI distances were shorter in radiographs compared to cadaveric measurements, the CI-TM distance was longer in radiological images. This finding reinforces the importance of using accurate anatomical measurements when planning minimally invasive surgeries. Furthermore, angular measurements at anatomical landmarks such as the SIAS, TM, and CI showed significant differences between cadaveric and radiological measurements. In cadavers, the average angle at the SIAS was 121.27° on the right side, whereas in radiological assessments, the angle was 130.38°. These angular data provide more detailed information about the spatial configurations of bony structures, going beyond mere linear morphometric comparisons. Particularly in orthopedic surgical planning, applications such as implant placement, guidance of incision lines, and positioning of prosthetic components benefit from these angular measurements by providing surgeons with a more comprehensive anatomical insight. Moreover, identifying angular inconsistencies between different measurement methods (cadaveric vs. radiographic) allows for questioning which methods are more reliable in clinical practice. Also, including angular measurements alongside linear dimensions in this study contributed to a more holistic understanding of the three-dimensional relationships of anatomical structures and enabled strengthening of educational and clinical inferences applicable to practice.

The discrepancies observed between cadaveric and radiological measurements are best understood as reflections of inherent methodological differences rather than measurement errors. Cadaveric assessments, conducted through direct physical manipulation, are susceptible to postmortem tissue changes, loss of soft tissue elasticity, and potential distortion during dissection. Conversely, radiological measurements especially those obtained via PACS benefit from high-resolution imaging, digital precision, and standardized positioning that better reflects in vivo anatomy. These technical advantages enable more consistent identification of bony landmarks in radiographs. As such, the differences between the two modalities underscore their complementary strengths. Integrating both approaches can enhance the accuracy, depth, and contextual relevance of morphometric analyses, ultimately contributing to more robust anatomical and clinical insights.

On the other hand, cadaver-based education has

become an essential component of modern surgical training, offering high-fidelity simulations that closely mimic real-life anatomical structures. This approach allows trainees to practice complex procedures such as hip arthroplasty in a safe and controlled environment, improving both technical skills and surgical confidence. Systematic reviews and participant feedback consistently emphasize its short-term effectiveness in skill acquisition and its perceived realism compared to live surgery [18-20]. For instance, hands-on cadaveric models have demonstrated clear benefits in procedures like transobturator tape surgery and trauma interventions, confirming their value in surgical skill development [21]. Despite limitations such as ethical concerns, high costs, and limited specimen availability, the consensus remains that cadaveric simulation bridges the gap between theoretical instruction and clinical practice, enhancing surgical competence and patient safety [18]. In this context, the anatomical measurements and visuals presented in our study offer practical value. By providing accurate reference points and morphometric data, these materials can further support cadaver-based training, contributing to the refinement of surgical techniques and the standardization of training protocols.

Limitations

One of the primary limitations of this study is the absence of radiological evaluation of cadaveric specimens. Ideally, the findings should have been correlated with radiological assessments conducted directly on cadavers to enhance the anatomical accuracy and clinical relevance of the results. However, due to institutional and administrative constraints, it was not feasible to transport cadavers to a hospital-based radiology center, where imaging is typically performed for clinical purposes only. As a result, radiological imaging of cadaveric specimens could not be incorporated into the methodology, which may have limited the comprehensive interpretation of morphometric findings. Additionally, the study was conducted on a limited number of cadaveric samples, which may restrict the generalizability of the results. Future studies involving a larger cadaver sample size are recommended to improve the statistical power and allow for broader anatomical variability to be assessed more accurately. Another limitation is the small number of female cadavers, which precluded sex-based comparisons. The unequal sex distribution in the cadaver sample (11 males, 3 females) posed a limitation in performing statistically valid comparisons between male and female specimens. Due to the small number of female cadavers, sex-specific analyses were not conducted. Instead, radiographic and cadaveric measurements were analyzed collectively, regardless of sex, and general mean values were calculated. This approach was adopted to ensure statistical feasibility and internal consistency. However, the lack of sex-stratified analysis may reduce the applicability of the findings to sex-specific anatomical variations. Therefore, this limitation should be acknowledged, and future studies should aim for a more balanced sex distribution to enable comparative analyses between male and female anatomy.

CONCLUSION

The findings of this study make a significant and original contribution by establishing standardized morphometric measurement protocols for key reference points in hip anatomy, based on data obtained through both cadaveric and radiological methods. To our knowledge, this is the first study to quantitatively compare morphometric measurements obtained through both cadaveric and radiological methods within the context of hip anatomy.

Our results support the hypothesis that standardized morphometric measurements enhance consistency and reliability in surgical applications. The first hypothesis of our study, stating that there are significant differences between Kirschner wire measurements on cadavers and anthropometric measurements from pelvic radiographs, was supported by our findings. The second hypothesis, proposing that Kirschner wire measurements before and after dissection are similar, was also confirmed. These results highlight the impact of measurement method on morphometric values while demonstrating that dissection does not significantly alter the spatial relationships of anatomical reference points. Moreover, the cadaveric measurements and anatomical visuals generated in this study provide students and residents with invaluable hands-on learning opportunities that reinforce theoret-

ical knowledge through practical experience, thereby boosting surgical confidence and competence in a controlled and safe environment. However, limitations such as the inability to perform radiological imaging directly on cadaveric specimens and the relatively small sample size, especially the limited number of female cadavers, may restrict the generalizability of our findings and necessitate cautious interpretation.

In conclusion, this study underscores the importance of integrating both cadaveric and radiological data for precise and safe surgical outcomes. The established morphometric standards have the potential to improve the accuracy of minimally invasive hip procedures and enhance the overall quality of hip arthroplasty surgeries. These findings lay important groundwork for future research aimed at refining anatomical reference points and advancing orthopedic surgical practices, ultimately contributing to improved patient safety and surgical success.

Ethical Statement

Ethical approval for the use of cadavers was granted by the relevant institutional authorities, and the entire study was approved by the Non-Invasive Clinical Research Ethics Committee of Cukurova University Faculty of Medicine on November 8, 2024 (Decision No: 149/7)

Data Availability

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

Authors' Contribution

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

Conflict of interest

The authors disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The authors disclosed that they did not receive any grant during the conduction or writing of this study.

Generative Artificial Intelligence Statement

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

Editor's note

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

REFERENCES

1. Huo MH, Parvizi J, Bal BS, Mont MA. What's new in total hip arthroplasty. J Bone Joint Surg Am. 2009;91(10):2522-2534. doi: 10.2106/JBJS.I.00801.
2. Bostan B, Sen C, Güneş T, Erdem M, Aytekin K, Erkorkmaz U. [Total hip arthroplasty using the anterolateral minimally invasive approach]. Acta Orthop Traumatol Turc. 2009;43(6):464-471. doi: 10.3944/AOTT.2009.464. [Article in Turkish]
3. Hochreiter J, Hejkrlik W, Emmanuel K, Hitzl W, Ortmaier R. Blood loss and transfusion rate in short stem hip arthroplasty. A comparative study. Int Orthop. 2017;41(7):1347-1353. doi: 10.1007/s00264-016-3365-2.
4. Zha GC, Zhu XR, Wang L, Li HW. Tranexamic acid reduces blood loss in primary total hip arthroplasty performed using the direct anterior approach: a one-center retrospective observational study. J Orthop Traumatol. 2022;23(1):12. doi: 10.1186/s10195-022-00638-7.
5. Zhai H, Huang J, Li L, et al. Deep learning-based workflow for hip joint morphometric parameter measurement from CT images. Phys Med Biol. 2023;68(22). doi: 10.1088/1361-6560/ad04aa.
6. Sugano N, Noble PC, Kamaric E, Salama JK, Ochi T, Tullos HS. The morphology of the femur in developmental dysplasia of the hip. J Bone Joint Surg Br. 1998;80(4):711-719. doi: 10.1302/0301-620x.80b4.8319.
7. Verma M, Joshi S, Tuli A, Raheja S, Jain P, Srivastava P. Morphometry of Proximal Femur in Indian Population. J Clin Diagn Res. 2017;11(2):AC01-AC04. doi: 10.7860/JCDR/2017/23955.9210.
8. Solomon LB, Howie DW, Henneberg M. The variability of the volume of os coxae and linear pelvic morphometry. Considera-

tions for total hip arthroplasty. J Arthroplasty. 2014;29(4):769-776. doi: 10.1016/j.arth.2013.08.015.
9. Zeckey C, Bogusch M, Borkovec M, et al. Radiographic cortical thickness parameters as predictors of rotational alignment in proximal femur fractures: A cadaveric study. J Orthop Res. 2019;37(1):69-76. doi: 10.1002/jor.24166.
10. Zhou X, Ji H, Guo J, Yang Y, Cai P, Zhang X. Modified osteotomy of posterolateral overhanging part of the trochanter via posterior approach for hip arthroplasty: an anatomical study. BMC Musculoskelet Disord. 2020;21(1):119. doi: 10.1186/s12891-020-3088-9.
11. Yee A, Padovano WM, Rowe AG, et al. The Effect of Surgical Video on Resident Performance of Carpal Tunnel Release: A Cadaveric Simulation-Based, Prospective, Randomized, Blinded Pilot Study. Plast Reconstr Surg. 2020;145(6):1455-1463. doi: 10.1097/PRS.0000000000006817.
12. Zucal I, Feder AL, Kyaw T, et al. An Innovative Simulation Model for Microvascular Training. Plast Reconstr Surg. 2022;150(1):189e-193e. doi: 10.1097/PRS.0000000000009209.
13. Zhang L, Huang T, Li C, et al. Race and Gender Differences in Anterior Cruciate Ligament Femoral Footprint Location and Orientation: A 3D-MRI Study. Orthop Surg. 2024;16(1):216-226. doi: 10.1111/os.13918.
14. Austin MS, Hozack WJ, Sharkey PF, Rothman RH. Stability and leg length equality in total hip arthroplasty. J Arthroplasty. 2003;18(3 Suppl 1):88-90. doi: 10.1054/arth.2003.50073.
15. Vanrusselt J, Vansevenant M, Vanderschueren G, Vanhoenacker F. Postoperative radiograph of the hip arthroplasty: what the radiologist should know. Insights Imaging. 2015;6(6):591-600. doi: 10.1007/s13244-015-0438-5.
16. Cantrell WA, Samuel LT, Sultan AA, Acuña AJ, Kamath AF. Operative Times Have Remained Stable for Total Hip Arthroplasty for >15 Years: Systematic Review of 630,675 Procedures. JB JS Open Access. 2019;4(4):e0047. doi: 10.2106/JBJS.OA.19.00047.
17. Heinz T, Vasilev H, Anderson PM, et al. The Direct Anterior Approach (DAA) as a Standard Approach for Total Hip Arthroplasty (THA) in Coxa Profunda and Protrusio Acetabuli? A Radiographic Analysis of 188 Cases. J Clin Med. 2023;12(12):3941. doi: 10.3390/jcm12123941.
18. Yiasemidou M, Gkaragkani E, Glassman D, Biyani CS. Cadaveric simulation: a review of reviews. Ir J Med Sci. 2018;187(3):827-833. doi: 10.1007/s11845-017-1704-y.
19. James HK, Chapman AW, Pattison GTR, Griffin DR, Fisher JD. Systematic review of the current status of cadaveric simulation for surgical training. Br J Surg. 2019;106(13):1726-1734. doi: 10.1002/bjs.11325.
20. Hughes JL, Katsogridakis E, Junaid-Siddiqi A, et al. The development of cadaveric simulation for core surgical trainees. Bull R Coll Surg Eng. 2019;101(1):38-43. doi: 10.1308/rcsbull.2019.38.
21. Selçuk İ, Tatar İ, Huri E. The effect of cadaveric hands-on training model on surgical skills and confidence for transobturator tape surgery. J Turk Ger Gynecol Assoc. 2019;20(4):243-246. doi: 10.4274/jtgga.galenos.2018.2018.0043.

The European Research Journal 2025;11(5):845-857 Original Article

DOI: https://doi.org/10.18621/eurj.1738116 *Emergency Medicine*

Comparison of five early warning scores in predicting mortality risk in patients presenting to the emergency department with acute dyspnea: qSOFA, NEWS2, MEWS, HASI, and SIL

Neslihan Ergün Süzer[1], Süleyman Alpar[2], Ali Cankut Tatlıparmak[3], Sarper Yılmaz[4]

[1]Department of Emergency Medicine, Kocaeli Darıca Farabi Training and Research Hospital, University of Health Sciences, Kocaeli, Türkiye; [2]Department of Emergency Medicine, Faculty of Medicine, İstanbul Beykent University, İstanbul, Türkiye; [3]Department of Emergency Medicine, Faculty of Medicine, Üsküdar University, İstanbul, Türkiye; [4]Department of Emergency Medicine, Medicana Zincirlikuyu Hospital, İstanbul, Türkiye

ABSTRACT

Objectives: This study aimed to compare five early warning scores - Quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score 2 (NEWS2), Modified Early Warning Score (MEWS), Hospital Alert Severity Index (HASI), and Shock Index-Lactate (SIL) - in predicting 30-day mortality in elderly patients presenting to the emergency department (ED) with acute dyspnea.

Methods: This was a single-center, retrospective observational study. A total of 764 patients aged 65 years or older presenting to the emergency department with acute dyspnea over a five-year period were included in this study. The predictive accuracy of each score was evaluated using AUROC analysis and logistic regression.

Results: Our findings demonstrated that the qSOFA score had the highest accuracy in predicting 30-day mortality (AUROC: 0.768). Among these scores, qSOFA showed the best performance in predicting mortality with a sensitivity of 72.9% and specificity of 74.6%. In logistic regression analysis, the qSOFA score demonstrated the strongest independent association with 30-day mortality (odds ratio [OR]: 5.23, P<0.001). The SIL score also showed a significant association with mortality (OR: 1.29, P=0.035). However, the HASI (P=0.092), MEWS (P=0.726), and NEWS2 (P=0.344) scores were not independently significant in multivariable analysis. Regarding mortality timing, qSOFA was identified as the most robust predictor for early death (within the first 3 days) with an AUROC of 0.801. It also demonstrated superior performance in predicting late in-hospital death (after 3 days) with an AUROC of 0.632 and post-discharge mortality within 30 days with an AUROC of 0.788. Other scores (HASI, MEWS, NEWS2, SIL) demonstrated lower performance in predicting mortality across different time intervals.

Conclusions: qSOFA demonstrated the most consistent and accurate performance among the evaluated scores. It may serve as a practical tool for early risk stratification in elderly patients with acute dyspnea in ED settings.

Keywords: Dyspnea, elderly, emergency department, qSOFA, early warning score

Received: July 8, 2025 **Accepted:** August 11, 2025 **Available Online:** August 15, 2025 **Published:** September 4, 2025

How to cite this article: Ergün Süzer N, Alpar S, Tatlıparmak AC, Yılmaz S. Comparison of five early warning scores in predicting mortality risk in patients presenting to the emergency department with acute dyspnea: qSOFA, NEWS2, MEWS, HASI, and SIL. Eur Res J. 2025;11(5):845-857. doi: 10.18621/eurj.1738116

Corresponding author: Sarper Yılmaz, MD., Assoc. Prof., Phone: +90 212 970 79 70, E-mail: sarperyilmaz08@gmail.com

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

Acute dyspnea is a prevalent and concerning symptom in emergency medicine, being one of the most frequent reasons for presentation to the emergency department (ED). In the United States alone, dyspnea accounts for approximately three to four million ED visits annually. Similarly, in the Asia-Pacific region, it constitutes around 5% of all ED presentations [1, 2]. In the ED setting, dyspnea is not only associated with high rates of admission to both general wards and intensive care units (ICU) but also serves as an independent predictor of poor clinical outcomes, including hospital readmissions, intubation, and increased mortality [3-5]. This high morbidity and mortality risk necessitates prompt diagnosis and efficient management by emergency physicians (EPs) [6].

For EPs, the primary goal when managing patients presenting with acute dyspnea in the ED is to rapidly optimize arterial oxygenation, assess the need for emergency airway management and respiratory support, and promptly identify life-threatening conditions[7, 8]. Additionally, it is essential to establish the most likely etiology of dyspnea, initiate appropriate treatment, and stabilize the patient if they are critically ill [7, 9-12]. However, achieving these objectives poses a significant challenge, given the often-limited clinical data available upon initial assessment. Accurate and timely triage and management decisions are crucial for improving patient outcomes, particularly in the elderly, who represent a significant proportion of patients presenting with acute dyspnea.

Epidemiological data indicate that the demographic profile of ED visits is shifting, with a growing proportion of elderly patients. In 2022, approximately one in ten ED visits involved patients aged 65 years or older, a figure projected to rise to one in six by 2050 [13]. Age-related deterioration in pulmonary and cardiovascular function complicates dyspnea management in the elderly [14]. Age-related physiological changes can obscure the clinical presentation, affecting physical examination findings, vital signs, and laboratory parameters, thereby increasing diagnostic complexity [15, 16].

Given the diagnostic challenges associated with acute dyspnea, particularly among older adults, the use of clinical and biochemical scoring systems has been advocated to aid EPs in risk stratification and management decisions. Several scoring systems, including the quick Sequential Organ Failure Assessment (qSOFA), National Early Warning Score 2 (NEWS2), Modified Early Warning Score (MEWS), Hospital Alert Severity Index (HASI), and the Shock Index-Lactate (SIL) score, have been evaluated for their prognostic utility in predicting mortality in critically ill patients presenting to the emergency department [17-19]. However, evidence remains limited regarding the comparative effectiveness of these scoring systems, particularly in elderly populations in acute dyspnea case.

This study aims to evaluate the prognostic accuracy of these five scoring systems in predicting 30-day mortality among elderly patients presenting to the ED with acute dyspnea.

METHODS

Study Design and Setting

This study was a retrospective, single-center observational analysis conducted in the ED of a tertiary care hospital. This study was approved by the Memorial Şişli Hospital Ethics Committee (Decision No: 004, Date: 26.12.2024). The primary objective of this study was to compare the prognostic accuracy of five commonly used risk assessment and decision-making tools (qSOFA, NEWS2, MEWS, HASI, and SIL) in predicting 30-day mortality among elderly patients presenting to the ED with acute dyspnea. The study specifically focused on assessing early death, late in-hospital death, and post-discharge death, aiming to determine the most reliable scoring system for mortality prediction in this vulnerable population.

Definition of Acute Dyspnea

Acute dyspnea was defined as "a subjective experience of breathing discomfort that consists of qualitatively distinct sensations that vary in intensity", in accordance with the consensus statement by the American Thoracic Society [20]. Acute dyspnea was defined as the sudden onset or worsening of breathing difficulty within the past 96 hours, based on institutional clinical protocols, differentiating it from chronic dyspnea, which persists for more than four to eight weeks. Patients with known chronic dyspnea (e.g., Chronic obstructive pulmonary disease [COPD], Congestive heart failure) were included only if there was a documented acute deterioration in symptoms within this period, accompanied by clinical signs of respira-

tory distress. Patients with missing data essential for the calculation of any of the five scores or for outcome analysis were excluded from the study

Missing Data Management

Patients with missing data essential for the calculation of any of the five scores or for outcome analysis (e.g., 30-day mortality) were excluded from the study. No data imputation techniques were applied due to the retrospective nature of the study.

Study Population:

Patients aged 65 years and older who presented to the ED with a chief complaint of acute dyspnea between January 1, 2019, and December 31, 2024, were eligible for inclusion in the study, and a total of 764 patients met the inclusion criteria. Exclusion criteria included patients who had experienced dyspnea symptoms for more than four days prior to admission, those with incomplete or missing medical records, and patients whose dyspnea was attributable to non-respiratory causes, such as traumatic injuries. Additionally, patients who were transferred from other healthcare facilities and those with documented palliative care status or end-of-life directives that limited aggressive treatment were also excluded. Data for the study were retrospectively collected from the hospital's electronic medical record (EMR) system, encompassing relevant demographic, clinical, and laboratory variables.

The following variables were extracted:

(a) Demographic data: Age, sex;

(b) Vital signs on ED admission: Systolic blood pressure (SBP) and diastolic blood pressure (DBP), respiratory rate (RR), heart rate (HR), peripheral oxygen saturation (SpO_2), and temperature;

(c) Neurological status: Glasgow Coma Scale (GCS);

(d) Laboratory data: Lactate levels, arterial blood gas (ABG) analysis, serum creatinine, estimated glomerular filtration rate (eGFR), B-type natriuretic peptide (BNP), C-reactive protein (CRP), and partial pressure of carbon dioxide (pCO_2);

(e) Comorbidities: Congestive heart failure, chronic kidney disease, diabetes mellitus;

(f) Medication history: Use of anticoagulants, diuretics, and home oxygen therapy;

(g) Outcomes: 30-day mortality, early death (within 3 days), late in-hospital death (after 3 days but before discharge), and post-discharge death (within 30 days after discharge);

(h) Intervention data: Requirement for oxygen therapy, mechanical ventilation, intensive care unit (ICU) admission, and vasopressor support.

Risk Assessment Scores

Five commonly used scoring systems were calculated for each patient upon ED admission:

(1) qSOFA: Systolic blood pressure, respiratory rate, and altered mental status [21].

(2) NEWS2: Respiratory rate, oxygen saturation, supplemental oxygen, systolic blood pressure, heart rate, consciousness (ACVPU=Alert, Confusion, Verbal, Pain, and Unresponsive), temperature [22].

(3) MEWS: Systolic blood pressure, heart rate, respiratory rate, temperature, consciousness (AVPU) [23].

(4) HASI: The HASI score is calculated using the formula: SpO_2 / (age × shock index), where shock index = heart rate / systolic blood pressure [24].

(5) SIL: The SIL score is calculated as: serum lactate level × shock index where shock index = heart rate / systolic blood pressure [17].

Statistical Analysis

All statistical analyses were performed using R version 4.4.2 (R Foundation for Statistical Computing, Vienna, Austria). Continuous variables were summarized using mean ± standard deviation or median with interquartile ranges, based on visual assessment of distribution. Categorical variables were reported as counts and percentages. Between-group comparisons for continuous variables were conducted using Student's t-test or the Wilcoxon rank-sum test, as appropriate. Categorical variables were compared using Pearson's chi-squared test or Fisher's exact test when expected counts were low. To assess the prognostic accuracy of five scoring systems AUROC values with 95% confidence intervals were calculated. Optimal thresholds were identified using Youden's index, and corresponding sensitivity, specificity, and positive and negative likelihood ratios were reported. Pairwise comparisons between AUROCs were conducted using DeLong's test for correlated ROC curves. A multivariable logistic regression model was built to evaluate the independent association of each scoring system with 30-day mortality. All five scores were included in the

model simultaneously. Model performance was internally validated using 10-fold cross-validation on the training dataset (80% of the sample), and tested on a hold-out test dataset (20%). Discriminative performance was summarized using AUROC. Model calibration was assessed using the Hosmer-Lemeshow goodness-of-fit test, Brier score, calibration intercept and slope, Emax, and Eavg statistics. To investigate associations between scoring systems and the timing of mortality, a multinomial logistic regression model was fitted with "Survived" as the reference category. The dependent outcome included three mortality subtypes: early death, late in-hospital death, and post-discharge death. Adjusted odds ratios with 95% confidence intervals were reported. Additionally, AUROC values for each score were calculated using one-vs-rest binary ROC analyses for each outcome subtype.

Post-hoc power and effect size analyses were conducted to evaluate the adequacy of the sample. For the primary comparison between survivors and deceased patients, standardized mean differences (Cohen's d) were calculated for each score. The qSOFA score yielded a Cohen's d of 0.77, indicating a large effect size. The NEWS2, MEWS, and HASI scores demon-

Table 1. Baseline characteristics of patients stratified by 30-day mortality status

Variable	30-day survivors (n=657)	30-day deceased (n=107)	Mean difference OR (95% CI)	P value
Age (years)	76.6±7.4	81.0±7.8	4.35 (2.76-5.95)	**<0.001**
Female sex, n (%)	295 (44.9%)	57 (53.3%)		0.132
Systolic blood pressure (mm Hg)	146.3±29.2	128.1±33.7	18.15 (11.33-24.97)	**<0.001**
Diastolic blood pressure (mm Hg)	75.4±15.1	71.1±17.7	4.39 (0.84-7.94)	**0.017**
Heart rate (beats/min)	97.7± 1.2	99.6±23.4		0.420
Respiratory rate (breaths/min)	16.0 [16.0-18.0]	18.0 [16.0-22.0]		**<0.001**
Peripheral oxygen saturation (%)	93.0 [90.0-96.0]	91.0 [87.5-96.0]		0.134
Body temperature (°C)	36.3 [36.0-37.0]	36.3 [36.1-36.8]		0.705
Shock index	0.7 [0.6-0.8]	0.8 [0.6-1.0]		**<0.001**
GCS	15.0 [15.0-15.0]	14.0 [12.0-15.0]		**<0.001**
Lactate, mmol/L	1.6 [1.3-2.3]	2.1 [1.3-3.2]		**0.003**
Serum creatinine (mg/dL)	1.2 [1.0-1.5]	1.7 [1.4-2.2]		**<0.001**
eGFR (mL/min/1.73m²)	56.6±18.3	36.7±15.2	19.82 (16.61-23.02)	**<0.001**
BNP (pg/mL)	1189.0 [710.0-2502.0]	1957.0 [1021.0-3300.5]		**<0.001**
CRP (mg/L)	34.3 [22.2-50.2]	93.4 [52.5-162.4]		**<0.001**
Arterial pH	7.4 [7.3-7.4]	7.3 [7.2-7.3]		**<0.001**
Arterial pCO_2 (mm Hg)	42.1±7.9	49.8±12.0	7.76 (5.40-10.11)	**<0.001**
Urine output on arrival (mL/h)	49.6±14.8	25.1±9.3	24.44 (22.34-26.54)	**<0.001**
Charlson comorbidity index	4.0±2.1	6.0±2.7	1.97 (1.44-2.50)	**<0.001**

Data are shown as mean±standard deviation or median [interquartile range] or n (%). BNP=B-type natriuretic peptide, CRP=C-reactive protein; eGFR=estimated glomerular filtration rate; GCS=Glasgow Coma Scale, pCO_2=partial pressure of carbon dioxide, OR=odds ratio

strated moderate-to-large effect sizes (Cohen's d ranging from 0.45 to 0.70). For subgroup comparisons, the largest and smallest diagnostic categories were acute heart failure (n=218) and pulmonary embolism (n=41). Assuming a two-sided alpha of 0.05 and a power of 80%, the study had sufficient power to detect an absolute mortality difference of 20% between these groups.

RESULTS

A total of 764 elderly patients presenting to the emergency department with dyspnea were included in the final analysis. Of these, 107 (14.0%) died within 30 days, and 657 (86.0%) survived. Based on available discharge diagnoses, 218 patients (28.5%) were classified as having acute heart failure, 165 (21.6%) had pneumonia, 112 (14.7%) had COPD exacerbation, 41 (5.4%) had pulmonary embolism, and 86 (11.3%) had other or mixed causes. In 142 patients (18.5%), no definitive etiology could be identified. The corresponding 30-day mortality rates were 14.2%, 12.1%, 11.7%, 15.8%, 12.8%, and 13.4%, respectively. No subgroup showed a markedly different mortality profile (P>0.05). Table 1 presents the baseline characteristics stratified by 30-day mortality status. Deceased patients were older and exhibited lower blood pressures, lower urine output, higher lactate and creatinine levels, and greater comorbidity burden compared to survivors. Significant differences were also observed in mental status, respiratory rate, Glasgow Coma Scale

Table 2. Clinical status, risk scores, and outcomes by 30-day mortality status

Variable	30-day survivors (n=657)	30-day deceased (n=107)	Mean difference OR (95% CI)	P value
Clinical frailty scale	4.7±1.1	6.1±0.6	1.39 (1.25-1.53)	**<0.001**
HASI score	1.5±0.6	2.0±0.9	0.47 (0.31-0.64)	**<0.001**
NEWS2 score	3.1±2.3	4.8±2.8	1.77 (1.20-2.34)	**<0.001**
MEWS score	1.9±1.5	2.8±1.6	0.94 (0.62-1.27)	**<0.001**
qSOFA score	0.0 [0.0-1.0]	1.0 [0.0-2.0]		**<0.001**
SIL score	6.1±1.3	6.5±1.5	0.38 (0.07-0.69)	**0.017**
Congestive heart failure	242 (36.8%)	68 (63.6%)		**<0.001**
Chronic kidney disease	413 (62.9%)	101 (94.4%)		**<0.001**
Diabetes mellitus	221 (33.6%)	31 (29.0%)		**0.400**
Dependent mobility	202 (30.7%)	91 (85.0%)		**<0.001**
Delirium at presentation	74 (11.3%)	37 (34.6%)		**<0.001**
Home oxygen therapy	56 (8.5%)	22 (20.6%)		**<0.001**
Diuretic use	416 (63.3%)	85 (79.4%)		**0.002**
Anticoagulant use	165 (25.1%)	38 (35.5%)		**0.032**
Oxygen use in ED	47 (7.2%)	17 (15.9%)		**0.005**
Oxygen delivery (mechanical)	13 (2.0%)	25 (23.4%)		**<0.001**
ICU admission	191 (29.1%)	45 (42.1%)		**0.010**
Oliguria	57 (8.7%)	74 (69.2%)		**<0.001**
Discharged from ED	434 (66.1%)	26 (24.3%)		**<0.001**

Data are shown as mean±standard deviation or n (%). ED=Emergency department, HASI=Hospital Alert Severity Index, ICU=intensive care unit, MEWS=Modified Early Warning Score, NEWS2=National Early Warning Score 2, qSOFA=Quick Sequential Organ Failure Assessment, SIL=Shock Index-Lactate score, GCS=Glasgow Coma Scale, CI=Confidence interval, OR=odds ratio

Table 3. Performance metrics of scoring systems for predicting 30-day mortality

Score	AUROC (95% CI)	Cut-off	Sensitivity (95% CI)	Specificity (95% CI)	+LR OR (95% CI)	–LR OR (95% CI)
HASI	0.650 (0.594-0.707)	≥3.0	24.3% (16.5-33.5%)	97.3% (95.7-98.4%)	8.87 (5.04-15.61)	0.78 (0.70-0.87)
MEWS	0.671 (0.621-0.722)	≥2.0	75.7% (66.5-83.5%)	51.0% (47.1-54.9%)	1.54 (1.35-1.76)	0.48 (0.34-0.67)
qSOFA	0.768 (0.720-0.817)	≥1.0	72.9% (63.4-81.0%)	74.6% (71.1-77.9%)	2.87 (2.41-3.42)	0.36 (0.27-0.50)
SIL score	0.566 (0.503-0.628)	≥7.47	28.0% (19.8-37.5%)	87.4% (84.6-89.8%)	2.22 (1.54-3.19)	0.82 (0.73-0.93)
NEWS2	0.687 (0.631-0.743)	≥4.0	67.3% (57.5-76.0%)	65.3% (61.5-68.9%)	1.94 (1.64-2.30)	0.50 (0.38-0.66)

HASI=Hospital Alert Severity Index, MEWS=Modified Early Warning Score, qSOFA=Quick Sequential Organ Failure Assessment, SIL=Shock Index-Lactate, NEWS2=National Early Warning Score 2, AUROC=Area under the receiver operating characteristic curve; +LR=Positive likelihood ratio, –LR=Negative likelihood ratio, CI=Confidence interval, OR=odds ratio

scores, and key laboratory markers such as BNP, CRP, and pCO_2.

As summarized in Table 2, all five risk scores were significantly higher in the deceased group. The median [interquartile ranges] qSOFA score was 1.0 [0.0-2.0] in deceased patients versus 0.0 [0.0-1.0] among survivors (P<0.001). Similar differences were seen across HASI, MEWS, NEWS2, and SIL scores, as well as in frailty, mobility status, and the frequency of acute interventions.

Table 3 outlines the diagnostic performance of each scoring system for predicting 30-day mortality. The qSOFA score showed the highest discriminative ability with an AUROC of 0.768 (95% confidence interval [CI]: 0.720-0.817), followed by NEWS2 (0.687), MEWS (0.671), HASI (0.650), and SIL (0.566). Cut-off values based on the Youden index, corresponding sensitivities, specificities, and likelihood ratios are also provided.

Fig. 1 displays the ROC curves with 95% confidence bands for all scores. Pairwise comparisons using the DeLong test are shown in Table 4. qSOFA outperformed all other scores except NEWS2 (P=0.010). MEWS and NEWS2 did not differ significantly (P=0.520), while HASI and SIL demonstrated significantly lower AUROCs than qSOFA and MEWS.

In the multivariable logistic regression analysis including all five scoring systems, only the qSOFA score and the SIL score were independently associated with 30-day mortality. Prior to modeling, collinearity was assessed; all variance inflation factors (VIFs) ranged from 1.3 to 2.1, indicating low collinearity and supporting the simultaneous inclusion of all scores. The qSOFA score demonstrated the strongest independent association with mortality, with an adjusted odds ratio (OR) of 5.23 (95% CI, 3.30-8.47;P<0.001), followed by the SIL score with an OR of 1.29 (95% CI, 1.02-1.63; P=0.035). The HASI score showed a borderline association (OR, 1.60; 95% CI, 0.93-2.77; P=0.092), while MEWS and NEWS2 were not statistically significant predictors in the adjusted model (P=0.726 and P=0.344, respectively) (Table 5).

The logistic model demonstrated strong discriminative ability, with an AUROC of 0.808 in the training set (10-fold cross-validation) and 0.833 in the test set. Calibration was satisfactory: the Brier score was 0.084 in the test set, the Hosmer-Lemeshow test was non-significant (χ^2 = 6.79, P=0.559), and the calibration

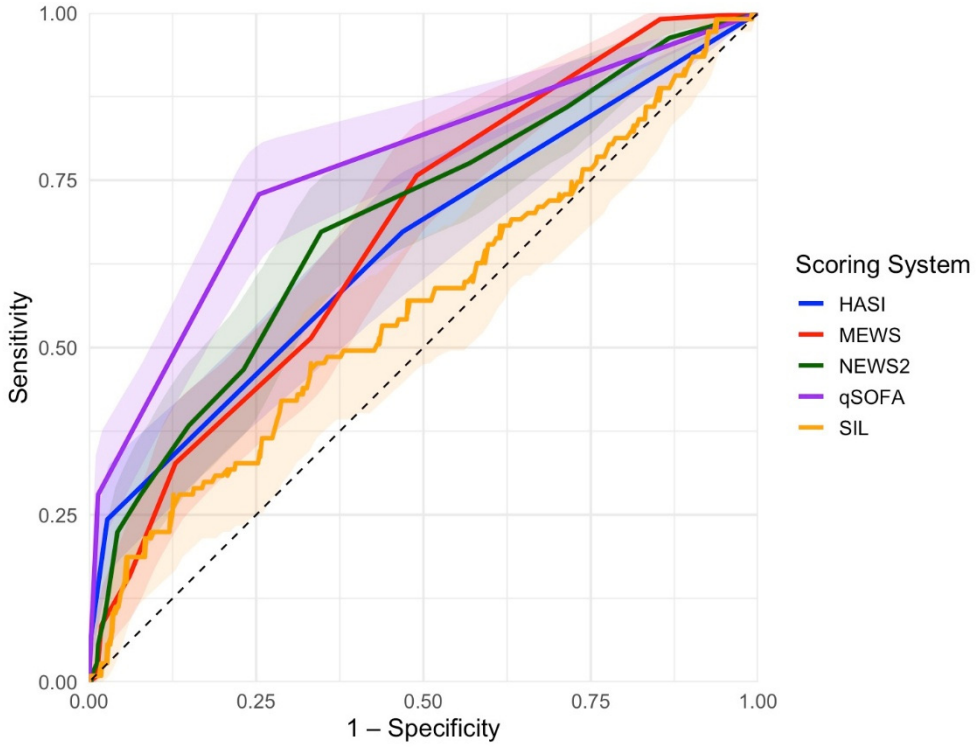

Fig. 1. Receiver Operating Characteristic (ROC) curves with 95% confidence interval for each scoring system.

slope and intercept were 1.22 and 0.32, respectively. Model calibration was assessed also using a LOESS-smoothed calibration curve comparing predicted probabilities with observed 30-day mortality. Visual inspection demonstrated good agreement with the ideal diagonal line, particularly in the low- and mid-risk ranges (Fig. 2). The model's maximum calibration error (Emax) was 0.263, with an average error (Eavg) of 0.025. The discrimination index (Dxy) was 0.666, and the Nagelkerke R^2 was 0.403, indicating moderate explanatory power.

To assess the relationship between each score and

Table 4. Pairwise comparisons of AUROC values for scoring systems (DeLong test)

Comparison	AUROC difference (95% CI)	P value
HASI vs MEWS	–0.021 (–0.089 to 0.047)	0.546
HASI vs qSOFA	–0.118 (–0.182 to –0.053)	**<0.001**
HASI vs SIL	0.085 (0.011 to 0.158)	**0.024**
HASI vs NEWS2	–0.036 (–0.080 to 0.008)	0.106
MEWS vs qSOFA	–0.097 (–0.154 to –0.040)	**<0.001**
MEWS vs SIL	0.105 (0.046 to 0.165)	**<0.001**
MEWS vs NEWS2	–0.015 (–0.062 to 0.031)	0.520
qSOFA vs SIL	0.203 (0.118 to 0.287)	<0.001
qSOFA vs NEWS2	0.082 (0.019 to 0.144)	**0.010**
SIL vs NEWS2	–0.121 (–0.182 to –0.060)	**<0.001**

AUROC=Area under the receiver operating characteristic curve, HASI=Hospital Alert Severity Index, MEWS=Modified Early Warning Score, qSOFA=Quick Sequential Organ Failure Assessment, SIL=Shock Index-Lactate, NEWS2=National Early Warning Score 2, CI=Confidence interval

Table 5. Independent association of each score with 30-day mortality in multivariable logistic regression

Scores	Adjusted Odds Ratio (95% CI)	P value
HASI	1.60 (0.93-2.77)	0.092
MEWS	1.05 (0.81-1.36)	0.726
qSOFA	5.23 (3.30-8.47)	**<0.001**
SIL score	1.29 (1.02-1.63)	**0.035**
NEWS2	0.91 (0.75-1.10)	0.344

HASI=Hospital Alert Severity Index, MEWS=Modified Early Warning Score, qSOFA=Quick Sequential Organ Failure Assessment, SIL=Shock Index-Lactate, NEWS20=National Early Warning Score 2, CI=Confidence interval

the timing of death, a multinomial logistic regression model was fitted using "Survived" as the reference category. Table 6 combines the adjusted odds ratios for early, late in-hospital, and post-discharge mortality for each score with their corresponding AUROC values based on one-vs-rest ROC analyses. The qSOFA score showed the strongest and most consistent predictive value across all mortality timing categories: early death (OR: 4.57, AUROC: 0.801), late in-hospital death (OR: 4.39, AUROC: 0.632), and post-discharge death (OR: 7.78, AUROC: 0.788). Other scores showed modest performance, and no additional score maintained independent significance across categories. To visualize these findings, Fig. 3 displays the distribution of survival outcomes across tertiles of each scoring system, allowing a visual assessment of how each score stratifies patients by mortality timing.

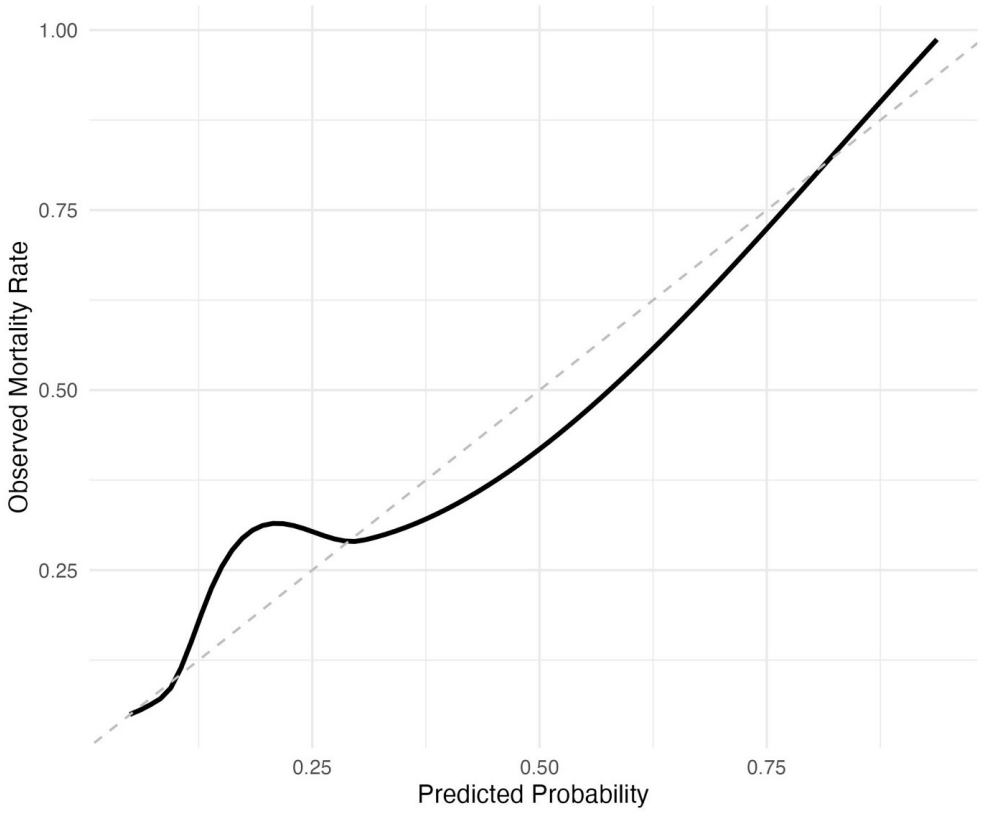

Fig. 2. Calibration plot of the multivariable logistic regression model including qSOFA, MEWS, NEWS2, HASI, and SIL scores.

Table 6. Multivariable logistic regression results and overall model performance metrics

Score	Early death OR (95% CI)	Early death AUROC (95% CI)	Late in-hosp death OR (95% CI)	Late in-hosp death AUROC (95% CI)	Post-discharge death OR (95% CI)	Post-discharge death AUROC (95% CI)
HASI	1.99 (0.87-4.54)	0.700 (0.574-0.825)	1.21 (0.55-2.67)	0.587 (0.504-0.669)	1.51 (0.83-2.76)	0.640 (0.557-0.724)
MEWS	1.41 (0.92-2.16)	0.727 (0.613-0.841)	0.67 (0.45-0.99)	0.505 (0.434-0.577)	1.11 (0.82-1.51)	0.718 (0.652-0.783)
qSOFA	4.57 (2.08-10.04)	0.801 (0.718-0.884)	4.39 (2.28-8.46)	0.632 (0.540-0.724)	7.78 (4.41-13.73)	0.788 (0.721-0.855)
SIL	1.00 (0.65-1.53)	0.561 (0.414-0.708)	1.30 (0.93-1.83)	0.521 (0.422-0.620)	1.23 (0.92-1.62)	0.583 (0.496-0.671)
NEWS2	0.96 (0.72-1.28)	0.741 (0.622-0.860)	1.01 (0.77-1.33)	0.576 (0.495-0.657)	0.95 (0.77-1.17)	0.696 (0.614-0.77

AUROC=Area under the receiver operating characteristic curve, HASI=Hospital Alert Severity Index, MEWS=Modified Early Warning Score, qSOFA=Quick Sequential Organ Failure Assessment, SIL=Shock Index-Lactate, NEWS2=National Early Warning Score 2, CI=Confidence interval, OR=odds ratio, hosp=hospital

DISCUSSION

In this study, we compared the prognostic values of five commonly used risk assessment and decision-making tools (qSOFA, NEWS2, MEWS, HASI, and SIL) for predicting mortality in elderly patients presenting to the ED with acute dyspnea. Among these tools, qSOFA demonstrated the highest performance in predicting 30-day mortality, as well as early death, late in-hospital death, and post-discharge death. qSOFA includes respiratory rate, systolic blood pressure, and mental status—three critical indicators of organ hypoperfusion and early sepsis, which are common pathways in acute dyspnea among elderly patients [25]. Conversely, composite scores like NEWS2 and MEWS incorporate variables such as temperature and supplemental oxygen use, which may not reflect immediate life-threatening derangements. This may reduce their discriminative power in high-risk elderly populations presenting with acute respiratory compromise. Similarly, SIL and HASI, though valuable in broader settings, rely on metrics (e.g., lactate, age, shock index) that may not respond as dynamically in early deterioration phases.

Compared to the other four scores, qSOFA emerged as the most effective tool in identifying elderly patients at risk of rapid deterioration during the hospital stay, as well as those who may experience worsening even after discharge. The key feature that makes qSOFA superior is its inclusion of low blood pressure, increased respiratory rate, and altered mental status, which are commonly encountered in elderly patients [26]. These parameters are closely associated with mortality in elderly patients presenting with dyspnea. Therefore, the use of qSOFA in the ED setting has proven to be highly valuable for rapid risk assessment and identifying patients requiring urgent intervention[27]. Recent literature on the use of qSOFA in the ED suggests that it outperforms other risk assessment tools in predicting mortality across various critical conditions, including trauma, pneumonia, sepsis, and infections [28-31]. Given the intersection of our study cohort - elderly patients with acute dyspnea presenting to the ED - we observed that the qSOFA score successfully identified mortality risk, proving to be a practical, cost-effective, and efficient tool for patient management. In contrast, the relatively lower performance of NEWS2 and MEWS may result from their re-

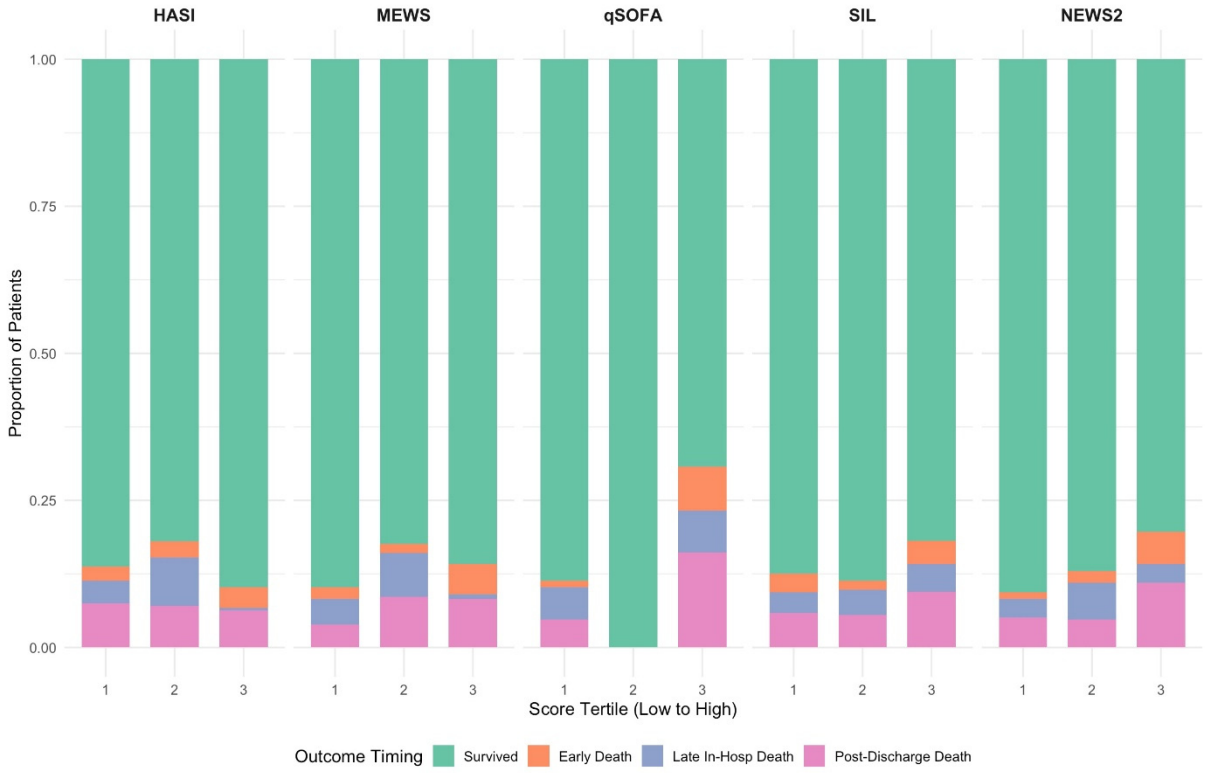

Fig. 3. Distribution of mortality timing across tertiles of each scoring system.

liance on parameters that may be chronically altered in the elderly, such as heart rate or baseline oxygen saturation, potentially limiting their discriminative power. Similarly, the HASI and SIL scores incorporate indices such as the shock index and lactate levels, which may fluctuate due to age-related physiological changes or chronic comorbidities, reducing their specificity in this patient group.

Our study revealed that SIL and HASI scores were inadequate for predicting mortality in elderly patients presenting with acute dyspnea. One of the main components of the SIL score, lactate level, can be elevated in conditions such as heart failure or chronic respiratory failure, even without sepsis or shock [32, 33]. This may lead to an inaccurate reflection of hemodynamic instability in elderly patients, thus compromising the reliability of the SIL score in this population. Similarly, the HASI score did not perform adequately in predicting mortality. The lower prognostic power of HASI, particularly in complex clinical scenarios such as acute dyspnea, limits its applicability in elderly patients [34]. Increased comorbidities and hemodynamic fluctuations in the elderly population may have contributed to HASI's limited predictive accuracy.

NEWS2 and MEWS scores also showed insufficient performance in predicting mortality related to acute dyspnea in elderly patients. The single parameters used in these scores may not adequately capture the multifactorial causes of dyspnea in the elderly, thereby limiting their prognostic accuracy. Frequent physiological changes and comorbid conditions encountered in elderly patients further restrict the utility of these scoring systems in clinical decision-making processes [35]. Therefore, when using NEWS2 and MEWS for managing acute dyspnea, it is crucial to take into account the unique clinical characteristics of elderly patients.

From a clinical perspective, the findings of this study suggest that incorporating qSOFA into the routine evaluation of elderly patients with acute dyspnea can enhance early identification of individuals at high risk of mortality. This is particularly relevant in resource-limited ED, where rapid and cost-effective tools are essential for prioritizing care. The simplicity of qSOFA - relying solely on systolic blood pressure, respiratory rate, and mental status - makes it a practical

triage tool even in settings with limited access to laboratory or imaging resources. By enabling faster decisions regarding the need for ICU admission, escalation of care, or early discharge planning, qSOFA can serve as a valuable aid in streamlining clinical workflows and improving patient outcomes in strained healthcare environments.

This study contributes to the existing literature by specifically focusing on elderly patients with acute dyspnea - a subgroup with unique physiological vulnerabilities and complex comorbid profiles often underrepresented in emergency risk stratification research. By evaluating multiple scoring systems in this context, our findings emphasize the utility of qSOFA as a simple, rapid, and cost-effective tool that can support clinical decision-making, particularly in emergency departments with limited resources. The results suggest that qSOFA may aid physicians in early identification of high-risk elderly patients, facilitate timely interventions, and optimize ICU triage pathways.

In summary, our study demonstrated that the qSOFA score is a robust tool for predicting mortality risk in elderly patients presenting to the ED with acute dyspnea. In particular, qSOFA showed clear superiority over other Early Warning Scores (EWS) in predicting critical outcomes, such as early death. In contrast, the SIL score, which relies heavily on lactate, proved to be less effective in this patient group. HASI, NEWS2, and MEWS also showed lower performance among elderly patients. Based on these results, it may be recommended to use qSOFA for rapid risk assessment and decision-making processes in elderly patients presenting with acute dyspnea in the ED. Performing quick and accurate risk stratification in this population can play a vital role in reducing mortality.

Limitations

This study has several limitations: (1) The study was conducted in a single tertiary care hospital and designed retrospectively, which may limit the generalizability of the findings; (2) The study did not stratify patients by the underlying etiology of dyspnea (e.g., acute heart failure, COPD exacerbation, pulmonary embolism), which may have influenced the performance of the evaluated scoring systems. Future research should consider diagnostic subgroup analyses to enhance the specificity and clinical applicability of early warning scores in heterogeneous dyspnea populations; (3) We also acknowledge the potential risk of missing data due to the retrospective nature of the study. To mitigate this, patients with incomplete data essential for score calculation or outcome assessment were excluded. Moreover, 30-day mortality follow-up was conducted using hospital electronic medical records, which are integrated with national death notification systems to capture out-of-hospital deaths as well; (4) Mortality data in the study were directly obtained from medical records, introducing the potential for subjective interpretation, which may affect the results; and (5) The study only assessed 30-day mortality; therefore, longer follow-up data are needed to provide more comprehensive insights into long-term survival and prognosis.

While our study underscores the prognostic utility of early warning scores - particularly Qsofa - in elderly patients presenting with acute dyspnea, its findings also have implications for improving clinical decision-making processes in emergency departments. Integrating validated scoring systems such as qSOFA into routine triage may support early identification of high-risk patients, allowing emergency physicians to prioritize resource allocation, determine the need for ICU admission, and initiate timely interventions. However, to enhance the external validity and generalizability of these results, future studies should adopt prospective, multicenter designs involving broader populations across different healthcare systems. Moreover, incorporating longer follow-up periods could provide additional insight into long-term mortality and morbidity outcomes, thereby strengthening the evidence base for integrating these tools into structured clinical algorithms and institutional protocols.

CONCLUSION

Among elderly patients presenting to the ED with acute dyspnea, qSOFA demonstrated the highest predictive accuracy for 30-day mortality compared to SIL, HASI, NEWS2, and MEWS. While it showed strong performance in identifying patients at risk of rapid deterioration and post-discharge events, further validation in more homogeneous diagnostic subgroups is warranted.

Ethics Approval and Consent to Participate

This study was approved by the Memorial Şişli

Hospital Ethics Committee (Decision No: 2024/004; date: 26.12.2024). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. The retrospective observational study included patients who presented to the EDs of two tertiary care hospitals: istanbul Beykent University Hospital and Memorial Bahçelievler Hospital. Due to the retrospective design of the study, the institutional review board granted a waiver of informed consent.

Data Availability

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

Authors' Contribution

Study Conception: NES, ACT, SY; Study Design: NES, ACT, SA; Supervision: ACT, SY; SA; Funding: N/A; Materials: NES, ACT, SY; SA; Data Collection and/or Processing: NES, ACT; Statistical Analysis and/or Data Interpretation: NES, ACT, SA; Literature Review: NES, ACT, SY, SA; Manuscript Preparation: NES, ACT, SY, SA; and Critical Review: ACT, SY, NES, SA.

Conflict of Interest

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

Financing

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

Acknowledgments

We would like to extend our gratitude to the individuals and institutions who contributed to this research.

Generative Artificial Intelligence Statement

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

Editor's Note

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

REFERENCES

1. Laribi S, Keijzers G, van Meer O, et al; AANZDEM and EURODEM study groups. Epidemiology of patients presenting with dyspnea to emergency departments in Europe and the Asia-Pacific region. Eur J Emerg Med. 2019;26(5):345-349. doi: 10.1097/MEJ.0000000000000571.
2. Kelly AM, Keijzers G, Klim S, et al; AANZDEM Study Group. An Observational Study of Dyspnea in Emergency Departments: The Asia, Australia, and New Zealand Dyspnea in Emergency Departments Study (AANZDEM). Acad Emerg Med. 2017;24(3):328-336. doi: 10.1111/acem.13118.
3. Pesola GR, Ahsan H. Dyspnea as an independent predictor of mortality. Clin Respir J. 2016;10(2):142-52. doi: 10.1111/crj.12191.
4. Demoule A, Baptiste A, Thille AW, et al; from the REVA Network (Research Network in Mechanical Ventilation). Dyspnea is severe and associated with a higher intubation rate in de novo acute hypoxemic respiratory failure. Crit Care. 2024;28(1):174. doi: 10.1186/s13054-024-04903-5.
5. Dupuis-Lozeron E, Soccal PM, Janssens JP, Similowski T, Adler D. Severe Dyspnea Is an Independent Predictor of Readmission or Death in COPD Patients Surviving Acute Hypercapnic Respiratory Failure in the ICU. Front Med (Lausanne). 2018;5:163. doi: 10.3389/fmed.2018.00163.
6. Collins S, Storrow AB, Kirk JD, Pang PS, Diercks DB, Gheorghiade M. Beyond pulmonary edema: diagnostic, risk stratification, and treatment challenges of acute heart failure management in the emergency department. Ann Emerg Med. 2008;51(1):45-57. doi: 10.1016/j.annemergmed.2007.07.007.
7. Baggish AL, Lloyd-Jones DM, Blatt J, et al. A clinical and biochemical score for mortality prediction in patients with acute dyspnoea: derivation, validation and incorporation into a bedside programme. Heart. 2008;94(8):1032-1037. doi: 10.1136/hrt.2007.128132.
8. Nazerian P, Vanni S, Volpicelli G, et al. Accuracy of point-of-care multiorgan ultrasonography for the diagnosis of pulmonary embolism. Chest. 2014;145(5):950-957. doi: 10.1378/chest.13-1087.
9. Januzzi JL Jr, Sakhuja R, O'donoghue M, et al. Utility of amino-terminal pro-brain natriuretic peptide testing for prediction of 1-year mortality in patients with dyspnea treated in the emergency department. Arch Intern Med. 2006;166(3):315-320. doi: 10.1001/archinte.166.3.315.
10. Baggish AL, Siebert U, Lainchbury JG, et al. A validated clinical and biochemical score for the diagnosis of acute heart failure: the ProBNP Investigation of Dyspnea in the Emergency Depart-

ment (PRIDE) Acute Heart Failure Score. Am Heart J. 2006;151(1):48-54. doi: 10.1016/j.ahj.2005.02.031.
11. Guttikonda SNR, Vadapalli K. Approach to undifferentiated dyspnea in emergency department: aids in rapid clinical decision-making. Int J Emerg Med. 2018;11(1):21. doi: 10.1186/s12245-018-0181-z.
12. Qaseem A, Etxeandia-Ikobaltzeta I, Mustafa RA, Kansagara D, Fitterman N, Wilt TJ; Clinical Guidelines Committee of the American College of Physicians. Appropriate Use of Point-of-Care Ultrasonography in Patients With Acute Dyspnea in Emergency Department or Inpatient Settings: A Clinical Guideline From the American College of Physicians. Ann Intern Med. 2021;174(7):985-993. doi: 10.7326/M20-7844.
13. Australian Institute of Health and Welfare, issuing body. Emergency department care: Australian hospital statistics Canberra: Australian Institute of Health and Welfare, 2015. Accessed 14 July 2025. http://nla.gov.au/nla.obj-767401731.
14. Ray P, Birolleau S, Riou B. [Acute dyspnoea in elderly patients]. Rev Mal Respir. 2002;19(4):491-503. [Article in French].
15. Janssens JP, Pache JC, Nicod LP. Physiological changes in respiratory function associated with ageing. Eur Respir J. 1999;13(1):197-205. doi: 10.1034/j.1399-3003.1999.13a36.x.
16. Schneider JL, Rowe JH, Garcia-de-Alba C, Kim CF, Sharpe AH, Haigis MC. The aging lung: Physiology, disease, and immunity. Cell. 2021;184(8):1990-2019. doi: 10.1016/j.cell.2021.03.005.
17. Devia Jaramillo G, Erazo Guerrero L, Florez Zuñiga N, Martin Cuesta RM. Evaluating the Accuracy of the SIL Score for Predicting the Sepsis Mortality in Emergency Department Triages: A Comparative Analysis with NEWS and SOFA. J Clin Med. 2024;13(24):7787. doi: 10.3390/jcm13247787.
18. Hincapié-Osorno C, van Wijk RJ, Postma DF, et al. Validation of MEWS, NEWS, NEWS-2 and qSOFA for different infection foci at the emergency department, the acutelines cohort. Eur J Clin Microbiol Infect Dis. 2024;43(12):2441-2452. doi: 10.1007/s10096-024-04961-1.
19. Vargun P, Yilmaz S, Tatliparmak AC, Karcioglu O. Should lactate levels be combined with rapid emergency medicine scores (REMS) to predict outcomes of patients with dyspnea. Signa Vitae. 2023;19(5):161-168. doi: 10.22514/sv.2023.024.
20. Parshall MB, Schwartzstein RM, Adams L, et al; American Thoracic Society Committee on Dyspnea. An official American Thoracic Society statement: update on the mechanisms, assessment, and management of dyspnea. Am J Respir Crit Care Med. 2012;185(4):435-452. doi: 10.1164/rccm.201111-2042ST.
21. Seymour CW, Liu VX, Iwashyna TJ, et al. Assessment of Clinical Criteria for Sepsis: For the Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3). JAMA. 2016 Feb 23;315(8):762-774. doi: 10.1001/jama.2016.0288.
22. National Early Warning Score (NEWS) 2 [Internet]. [cited 2025 Jan 24]. Available from: https://www.rcp.ac.uk/improving-care/resources/national-early-warning-score-news-2/
23. Modified Early Warning Score (MEWS) [Internet]. [cited 2025 Feb 5]. Available from: https://reference.medscape.com/calculator/461/modified-early-warning-score-mews
24. Chen PS, Hsieh CY, Jaw FS, Chen HK, Hsi KY, Chang HP. The hypoxia-age-shock index at triage is a useful and rapid tool. Am J Emerg Med. 2024;83:154-155. doi: 10.1016/j.ajem.2024.07.010.
25. López-Izquierdo R, Brio-Ibañez PD, Martín-Rodríguez F, et al. Role of qSOFA and SOFA Scoring Systems for Predicting In-Hospital Risk of Deterioration in the Emergency Department. Int J Environ Res Public Health. 2020;17(22):8367. doi: 10.3390/ijerph17228367.
26. Nishino T, Iimura K, Miura N. Elderly woman with disturbance of consciousness and respiratory failure. J Am Coll Emerg Physicians Open. 2020;1(4):658-659. doi: 10.1002/emp2.12084.
27. Güleç T, Yılmaz S, Ak R, Tatlıparmak AC, Karcıoğlu Ö. Can we recognize severe community-acquired pneumonia without pneumonia severity index? Use of modified qSOFA with procalcitonin. Heliyon. 2023;9(9):e19937. doi: 10.1016/j.heliyon.2023.e19937.
28. Meral G, Ardıç Ş, Günay S, et al. Comparative analysis of Glasgow Coma Scale, quick Sepsis-related Organ Failure Assessment, base excess, and lactate for mortality prediction in critically ill emergency department patients. Turk J Emerg Med. 2024;24(4):231-237. doi: 10.4103/tjem.tjem_45_24.
29. Shahsavarinia K, Moharramzadeh P, Arvanagi RJ, Mahmoodpoor A. qSOFA score for prediction of sepsis outcome in emergency department. Pak J Med Sci. 2020;36(4):668-672. doi: 10.12669/pjms.36.4.2031.
30. Kesselmeier M, Pletz MW, Blankenstein AL, et al. Validation of the qSOFA score compared to the CRB-65 score for risk prediction in community-acquired pneumonia. Clin Microbiol Infect. 2021;27(9):1345.e1-1345.e6. doi: 10.1016/j.cmi.2020.10.008.
31. Huang W, Yang P, Xu F, Chen D. Predictive value of qSOFA score for death in emergency department resuscitation room among adult trauma patients:a retrospective study. BMC Emerg Med. 2021;21(1):103. doi: 10.1186/s12873-021-00498-0.
32. Yuzefpolskaya M, Schwartz S, Ladanyi A, et al. The Role of Lactate Metabolism in Heart Failure and Cardiogenic Shock: Clinical Insights and Therapeutic Implications. J Card Fail. 2025:S1071-9164(25)00041-7. doi: 10.1016/j.cardfail.2025.01.011.
33. Tangpaisarn T, Drumheller BC, Daungjunchot R, Kotruchin P, Daorattanachai K, Phungoen P. Severe hyperlactatemia in the emergency department: clinical characteristics, etiology and mortality. BMC Emerg Med. 2024;24(1):150. doi: 10.1186/s12873-024-01071-1.
34. Hsieh CC, Liu CY, Tsai KC, Jaw FS, Chen J. The hypoxia-age-shock index at triage to predict the outcomes of Covid-19 patients. Am J Emerg Med. 2023;65:65-70. doi: 10.1016/j.ajem.2022.12.034.
35. Riccalton V, Threlfall L, Ananthakrishnan A, et al. Modifications to the National Early Warning Score 2: a Scoping Review. BMC Med. 2025;23(1):154. doi: 10.1186/s12916-025-03943-0.

The European Research Journal 2025;11(5):858-867 Original Article

DOI: https://doi.org/10.18621/eurj.1704067 *Medical Oncology*

Evaluating the role of HER2-low versus HER2-0 status in predicting response to neoadjuvant chemotherapy in hormone receptor-positive breast cancer

Alper Topal, Mustafa Başak

Department of Internal Medicine, Division of Medical Oncology, Faculty of Medicine, Tokat Gaziosmanpaşa University, Tokat, Türkiye.

ABSTRACT

Objectives: The classification of human epidermal growth factor receptor 2 (HER2)-low breast cancer has gained clinical relevance following the success of antibody-drug conjugates in this subgroup. However, its prognostic and predictive role, particularly in hormone receptor-positive (HR+) early breast cancer treated with neoadjuvant chemotherapy (NACT), remains unclear. This study aimed to evaluate the impact of HER2-low versus HER2-0 status on pathological complete response (pCR) and disease-free survival (DFS) in HR+ breast cancer patients undergoing NACT.

Methods: A total of 216 HR+ and HER2-negative early breast cancer patients treated with NACT at Tokat Gaziosmanpasa University Hospital between January 2014 and January 2024 were retrospectively analyzed. HER2-low was defined as IHC 1+ or 2+ without gene amplification by FISH. pCR was assessed via the Miller-Payne grading system. Survival analyses were conducted using the Kaplan-Meier method; multivariate analyses were performed using Cox regression.

Results: Of the 216 patients, 30 (13.9%) achieved pCR. There was no statistically significant difference in pCR (P=0.83) or DFS (P=0.12) between HER2-0 and HER2-low groups. However, patients with ER <10% had significantly higher pCR rates (P=0.005). Achieving pCR was associated with longer DFS (P=0.045).

Conclusions: HER2-0 and HER2-low subgroups exhibited similar responses to NACT in HR+ breast cancer. Low ER expression was independently associated with higher pCR. Larger prospective studies are warranted to further define the biological and clinical implications of HER2 expression levels in early-stage HR+ breast cancer.

Keywords: Neoadjuvant chemotherapy, HER2-low, breast cancer, pathologic complete response

The second leading cause of mortality for women and the most frequent type of cancer overall is breast cancer (BC) [1]. Consequently, a large proportion of patients are detected at an early stage thanks to early screening programmes in many countries [2]. Hormone receptor-positive (HR+) BC, defined by the expression of immunohistochemical positivity for the estrogen (ER) and/or progesterone (PR) receptor, is the most common subtype. It accounts for approximately 70% of all breast cancer patients [3, 4]. Neoadjuvant chemotherapy (NACT) is used primarily in individuals with biolog-

Received: May 22, 2025 **Accepted:** August 1, 2025 **Available Online:** August 6, 2025 **Published:** September 4, 2025

How to cite this article: Topal A, Başak M. Evaluating the role of HER2-low versus HER2-0 status in predicting response to neoadjuvant chemotherapy in hormone receptor-positive breast cancer. Eur Res J. 2025;11(5):858-867. doi: 10.18621/eurj.1704067

Corresponding author: Alper Topal, MD., Asisst. Prof., Phone: +90 356 252 16 16, E-mail: dralpertopal@gmail.com

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

ically aggressive tumours, including triple-negative breast cancer (TNBC), human epidermal growth factor receptor 2 (HER2)-positive BC, and in individuals having ER (+)/HER2 (-) BC diagnosed with high-risk clinicopathological features. NACT reduces the stage of the tumour and allows breast and axillary surgery to be reduced. It is increasingly being used to enable breast-conserving surgery (BCS) to be performed, thus avoiding mastectomy [5].

Although response to NACT is prognostic in all tumour types, TNBC and HER2 (+) BC had notably greater pathological complete response (pCR) rates than luminal subtypes [6]. pCR rates are much lower in ER (+)/HER2 (-) breast cancer. It is linked to a pathological full reaction, improved survival and lower recurrence rates [7]. Therefore, pCR is a surrogate marker.

In the light of information from anti-HER2 antibody-drug conjugate (trastuzumab deruxtecan) research studies, new subsets such as HER2-low and HER2-ultra-low have emerged [8-10]. There are studies in the literature suggesting that the biology, histological, and proliferative values of HER2-low and HER2-0 tumors vary [11, 12]. The effect of HER2-low/HER2-0 status in reaction to neoadjuvant treatment is of interest. In our research, our aim was to investigate the effect of HER2-0/HER2-lowistatus on pCR and disease-free survival (DFS) after NACT in HR (+) BC.

METHODS

Patient Selection

The data of patients who were identified and handled for BC at Tokat Gaziosmanpasa University Hospital between January 2014 and January 2024 were retrospectively evaluated. ER, PR, HER2 and Ki67 values were analysed immunohistochemically. Patients with HER2 immunohistochemistry (IHC) values of 0, 1+ and 2+ and no amplification of genes determined by fluorescent situ hybridisation (FISH) were considered HER2-negative. Individuals having HER2 values of 1+ and 2+ were categorised as 'HER2-low'. All patients underwent clinical staging with ultrasound and mammography preoperatively.

BC individuals determined as PRipositive or negative, ER positive, HER2 negative or low expression, and who received neoadjuvant chemotherapy (NACT) were analyzed in the research. Individuals having bilateral BC, male gender and distant metastatic disease were excluded. Variables such as tumour size, menopausal status, age, pathological axillary lymph node (ALN) count, clinical stage, receptor status, HER2 status, Ki67 level, and histological grade were analysed.

Neoadjuvant Chemotherapy (NACT)

A standard taxane and anthracycline-based chemotherapy regimen was applied in the study. The treatment protocol was four courses of doxorubicin (60 mg/m²) and cyclophosphamide (600 mg/m²) followed by weekly paclitaxel (80 mg/m²) or four courses of docetaxel (75-100 mg/m²) for 12 weeks. Postoperatively, all patients had adjuvant endocrine treatment, and all individuals who had BCS were advised to get adjuvant radiation.

Evaluation of NACT Response

Response to chemotherapy was evaluated with the routinely used Miller-Payne grading system. The Miller-Payne system is given in the following:

(1) Grade 1: No change or some alteration to individual malignant cells but no reduction in overall cellularity.

(2) Grade 2: A minor loss of tumour cells but overall cellularity still high; up to 30% loss.

(3) Grade 3: Between an estimated 30% and 90% reduction in tumour cells.

(4) Grade 4: A marked disappearance of tumour cells such that only small clusters or widely dispersed individual cell remain; more than 90% loss of tumour cells.

(5) Grade 5: No malignant cells identifiable in sections from the site of the tumour; only vascular fibroelastic stroma remains often containing macrophages. However, ductal carcinoma in situ (DCIS) may be present [12].

In the postoperative histopathological evaluation, tumour stage (ypT), lymph node stage (ypN), residual tumour size, surgical margins, number of removed and metastatic ALNs were analysed after NACT. pCR has been described as a lack of invasive tumor in breast tissue and metastasis in lymph nodes (ypT0, ypN0).

Statistical Analysis

SPSS 22.0 software was used for the statistical an-

alyze of the data (SPSSiInc., Chicago, Illinois). For comparative data, Fisher's exact test and chi-square tests were employed. Amongst the numerical parameters across two independent conditions, those with normal distribution were analysed by Student's t-test, and those without normal distribution were analysed by the Mann-Whitney U test. The univariate log-rank test was used to assess the impact of prognostic variables on pathological complete response. The 95% confidence interval (CI) was used to compute the hazard ratio (HR). The Cox proportional hazards model was used for multivariate analysis to assess the impact of prognostic variables on pathological complete response. To make an assessment regarding the prognostic variables influencing pathological complete response, both univariate and multivariate analyses were conducted using a logistic regression model. Survival studies were conducted using the Kaplan-Meier technique. DFS was defined as the interval between the first diagnosis and the death or recurrence of the disease. The significance level was set as ≤ 0.05.

Table 1. Data on pathological and clinical characteristics

			HER2 status		
		Total (n=216)	HER2-0 (n=63)	HER2-Low (n=153)	P value
Diagnosis age (year)	Median (min-max)	50 (28-87)			
	<50 years, n (%)	113 (52.3)	27 (42.9)	86 (56.2)	0.099
	≥50 years, n (%)	103 (47.6)	36 (57.1)	67 (43.8)	
Menopausal status, n (%)	Premenopausal	112 (51.8)	23 (36.5)	89 (58.2)	**0.004**
	Postmenopausal	104 (48.2)	40 (63.5)	64 (41.8)	
N stage, n (%)	NX and N0	18 (8.3)	7 (11.7)	11 (7.5)	0.414
	N1-N2-N3	189 (87.5)	53 (88.3)	136 (92.5)	
T stage, n (%)	TX and T1	62 (28.7)	19 (31.1)	43 (29.3)	0.868
	T2-T3-T4	146 (67.5)	42 (68.9)	104 (70.7)	
Clinical stage, n (%)	I	4 (1.8)	2 (3.4)	2 (1.4)	0.631
	II	125 (57.8)	35 (59.3)	90 (62.1)	
	III	75 (34.7)	22 (37.3)	53 (36.5)	
Grade (n=189), n (%)	Grade 1 and 2	154 (71.2)	43 (79.6)	111 (75.5)	0.579
	Grade 3	47 (21.7)	11 (20.4)	36 (24.5)	
ER Status	1-9%	5 (2.3)	3 (4.8)	2 (1.3)	0.307
	10-40%	11 (5.2)	3 (4.8)	8 (5.2)	
	>40%	200 (92.5)	57 (90.4)	143 (93.5)	
PR Status	Positive	186 (86.1)	55 (88.7)	131 (86.2)	0.823
	Negative	28 (12.9)	7 (11.3)	21 (13.8)	
Ki-67 (n=199), n (%)	<20 %	79 (36.5)	26 (45.6)	53 (37.3)	0.337
	≥20 %	120 (55.5)	31 (54.4)	89 (62.7)	
Surgery type, n (%)	Mastectomy	136 (62.9)	42 (66.7)	94 (61.4)	0.536
	Breast conserving	80 (37.1)	21 (33.3)	59 (38.6)	
pCR	Yes	30 (13.9)	8 (12.7)	22 (14.4)	0.831
	No	186 (86.1)	55 (87.3)	131 (85.6)	

Data are shown as median (minimum-maximum) or n (%). ER=estrogen receptor, PR=progesterone receptor, HER2=human epidermal growth factor receptor 2

RESULTS

The study included 216 patients. The median follow-up (mFU) time was 37 months. The median age of the participants was 50 years (min-max: 28-87). The number of premenopausal patients was 112 (51.8%). The number of patients with positive lymph nodes (N1-2-3) was 189 (87.5%). The clinical and pathological characteristics are presented in Table 1. There were 5(2.3%) patients with ER 1-9%, 11 patients (5.2%) with ER 10-40%, and 200 (92.5%) patients with ER over 40%. PR was negative in 28 (12.9%) patients.

Table 2. Clinicopathological characteristics and oncological outcome with respect to pathologic complete response achievement

			pCR achievement		
		Total (n=216)	Yes (n=30)	No (n=186)	P value
Diagnosis age (year)	<50 years	113	18 (15.9)	95 (84.1)	0.433
	≥50 years	103	12 (11.7)	91 (88.3)	
Menopausal status, n (%)	Premenopausal	112	17 (15.2)	95 (84.8)	0.694
	Postmenopausal	104	13 (12.5)	91 (87.5)	
N stage, n (%)	NX and N0	18	1 (5.6)	17 (94.4)	0.478
	N1-N2-N3	189	27 (14.4)	162 (85.7)	
T stage, n (%)	TX and T1	62	6 (9.7)	56 (90.3)	0.377
	T2-T3-T4	146	22 (15.1)	124 (84.9)	
Clinical stage, n (%)	I	4	0 (0)	4 (100)	0.651
	II	125	18 (18.4)	107 (85.6)	
	III	75	9 (12.0)	66 (88.0)	
Grade (n=189), n (%)	Grade 1 and 2	154	16 (10.4)	138 (89.6)	**0.029**
	Grade 3	47	11 (23.4)	36 (76.6)	
ER positivity, n (%)	1-9%	5	3 (60)	2 (40)	**0.001**
	10-40%	11	4 (36.4)	7 (63.6)	
	>40%	200	23 (11.5)	177 (88.5)	
PR status	Positive	186	27 (14.5)	159 (85.5)	0.385
	Negative	28	2 (7.1)	26 (92.9)	
HER2 score	HER2-0	63	8 (12.7)	55 (87.3)	0.831
	HER2-Low	153	22 (14.4)	131 (85.6)	
Ki-67 (n=212), n (%)	<20 %	79	5 (6.3)	74 (93.7)	**0.047**
	≥20 %	120	20 (16.7)	100 (83.3)	
Surgery type, n (%)	Mastectomy	136	19 (14.0)	117 (86.0)	0.99
	Breast conserving	80	11 (13.8)	69 (86.2)	
Relaps	Yes	34	2 (5.9)	32 (94.1)	0.182
	No	182	28 (15.4)	154 (84.6)	
DFS (months), median (SE, 95%CI)			NR	91.9 (81.81-101.95)	**0.045**

ER=estrogen receptor, PR=progesterone receptor, HER2=human epidermal growth factor receptor 2, pCR=pathologic complete response

One hundred and twenty patients (55.5%) had Ki67 ≥ 20% (Table 1).

Regarding the type of surgery performed, 136 (62.9%) patients underwent mastectomy and 80 (37.1%) patients underwent breast-conserving surgery. While 30 (13.9%) patients had pathological complete response, 186 (86.1%) patients did not achieve pCR.

When we looked at the distribution of clinical and pathological characteristics of HER2-0 and HER2-Low groups, only the number of premenopausal patients was not normally distributed (P=0.004). The rate of premenopausal patients was higher in the HER2-low group (58.2%).

According to the results of univariate analysis, tumour grade, ER and Ki67 percentages were seen to be the elements affecting pCR (P=0.029, P=0.001 and P=0.047, respectively). According to the results of multivariate analysis, only low ER receptor percentage was found to be a factor affecting pCR (P=0.005). Patients with ER receptor percentage less than 10% had 3.7 times more pCR (Table 2).

When analysing DFS according to pCR, the median DFS of pCR participants was not reached, whereas the median DFS of non-pCR patients was

Table 3. Variables influencing pathologic complete response

		Univariate		Multivariate	
		OR, 95%CI	P value	OR, 95% CI	P value
Diagnosis age (year)	<50 years	0.69 (0.31-1.52)	0.365		
	≥50 years				
Menopausal status	Premenopausal	0.79 (0.36-1.739	0.573		
	Postmenopausal				
N stage	NX and N0	2.83 (0.36-22.17)	0.321		
	N1-N2-N3				
T stage	TX and T1	1.65 (0.63-4.30)	0.30		
	T2-T3-T4				
Clinical stage	I	0.93 (0.42-2.06)	0.87		
	II				
	III				
Grade (n=189)	Grade 1 and 2	2.63 (1-12-6.17)	**0.02**	0.98 (0.32-2.99)	0.97
	Grade 3				
ER positivity	1-9%	0.27 (0.12-0.60)	**<0.001**	0.27 (0.11-1.67)	**0.005**
	10-40%				
	>40%				
PR Status	Positive	0.45 (0.10-2.02)	0.29		
	Negative				
HER2 score	Her2-0	0.86 (0.36-2.06)	0.74		
	Her2-Low				
Ki-67 (n=212)	≥20 %	2.96 (1.06-8.25)	**0.03**	2.48 (0.74-8.33)	0.13
	<20 %				
Surgery type	Mastectomy	0.98 (0.44-2.18)	0.96		
	Breast conserving				

ER=estrogen receptor, PR=progesterone receptor, HER2=human epidermal growth factor receptor 2, pCR=pathologic complete response, OR=odds ratio, CI=confidence interval

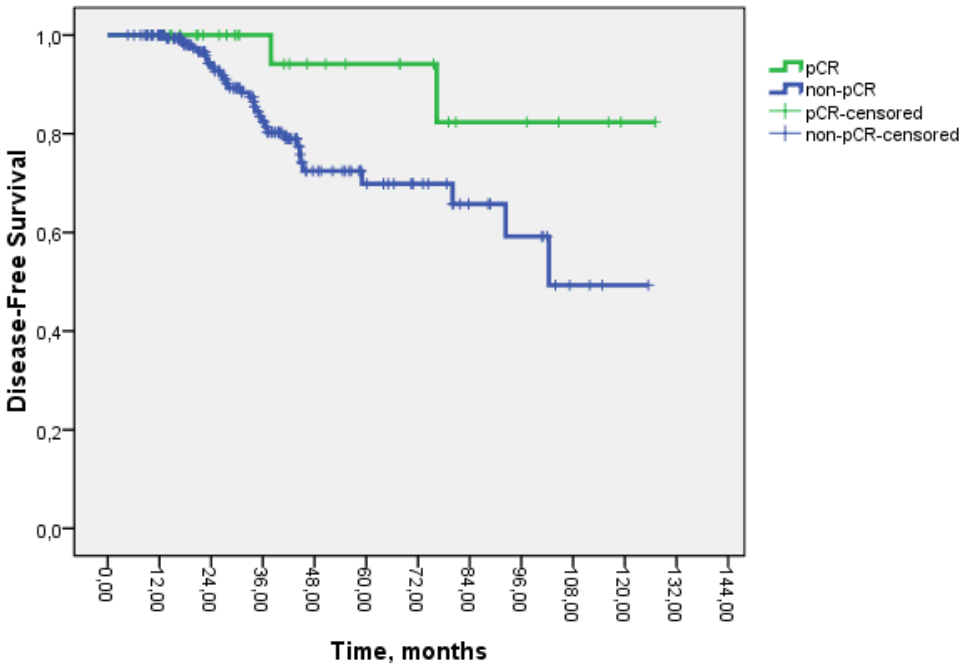

Fig. 1. **Associations between pathological complete response and disease-free survival.**

91.9 (81.81-101.95) months, which was statistically significant (P=0.045, Table 3 and Fig. 1).

There was no variation in pCR (P=0.83) or DFS (P=0.12) between the HER2-0 and HER2-low groups (Fig. 2). However, the 3 and 5-year DFS was numerically better in the HER2-0 group. The 3-year and 5-year DFS of HER2-0 patients were 88.7% and 86%, respectively, while those of HER2-Low patients were 83.2% and 69.6% (Table 4).

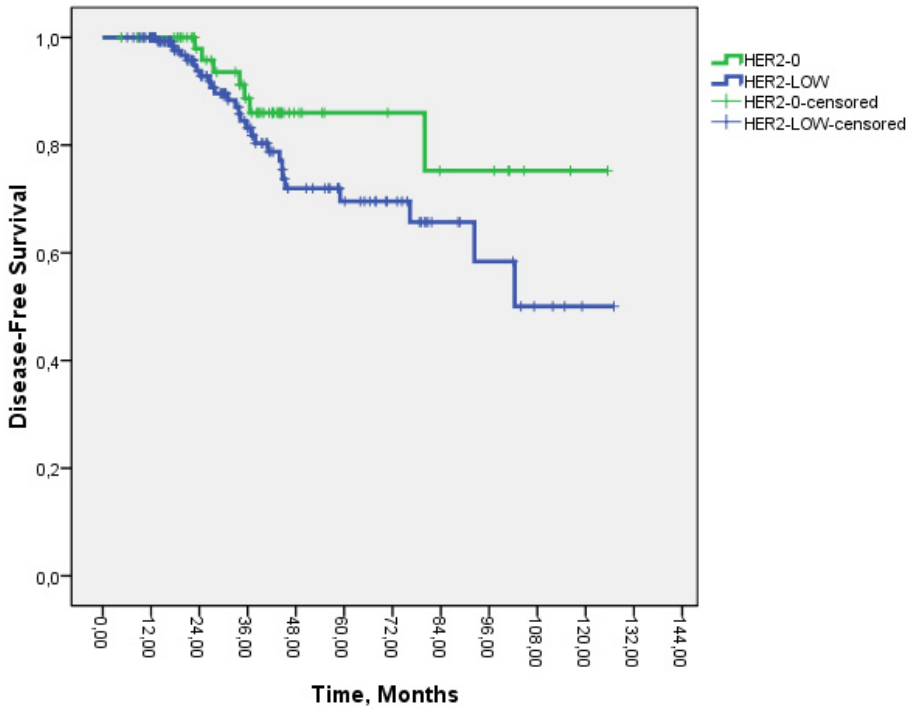

Fig. 2. **Associations between HER2 status and disease-free survival.**

DISCUSSION

HER2 low BC is a popular topic. Concern in HER2 low subgroups has grown after trastuzumab-deruxtecan's effectiveness in the HER2 low and extreme low subgroups was demonstrated [10]. We shared this interest and investigated the status influence of HER2 on complete response in HR (+) BC patients receiving NACT. It was shown that there was no varitation among HER2-0 and low subgroups regarding pCR and DFS. The DFS was longer for individuals who attained pCR than for those who could not. When the factors influencing pCR were analysed, low ER was found to be the factor influencing pCR.

pCR rates with neoadjuvant chemotherapy are lower in HR(+) tumours compared to HER2(+) and TNBC tumours [12, 13]. Complete response rates do not exceed 20% in the literature [12]. In our study, there was a pCR rate of 13% in accordance with the literature.

Zhou *et al*. [14] showed that HER2-0 and HER2-low groups had similar pCR values in a study of 325 patients receiving NACT. Similar results were found in another study of 855 patients conducted in Brazil [15]. In the meta-analysis of 2310 patients by Denkert *et al*. [16], the relationship between HER2-0 and HER2-low pCR in individuals with HR(+) and TNBC was analysed. It was seen that individuals with HER2-low in the general population and HR(+) had lower pCR rates. Baez-Navarro *et al*. [17] showed the same results in a study of 11721 individuals. de Moraes *et al*. [18] performed one of the largest meta-analyses on this topic. In this study, which included 70104 patients, it was shown that HER2-0 patients had a better pCR than HER2-low individuals. This was also true in the HR(+) subgroup. Although there are investigations in the literature showing no variation in pCR among HER2-0 and HER2-low patients [19, 20], large meta-analyses suggest that HER2-0 patients have better pCR rates. The fact that there was no difference in pCR among the two groups in our study may be explained by the relatively small size of our patient population.

It is known that HR (+) BC has a lower response rate to neoadjuvant treatment compared to TNBC [21, 22]. Dieci *et al*. [23] showed that ER-negative (<1%) and ER-low (1-9%) TNBC patients had similar pCR rates and emphasised that the ER-low group should be classified as TNBC. Similarly, another study showed that patients with ER-low and ER(-) TNBC had similar pCR rates [24]. Fuji *et al*. [25] showed in a study of 3055 patients that the pCR probability of patients with ER <10% tumours was notably greater than that of patients with ER≥10% tumours. In the study of PAM50 testing to predict NACT response, it was found that individuals with ER-low had TNBC like behaviour and should be treated with chemotherapy. It was shown that pCR rates were greater in individuals with ER-low [26]. Similarly, in our study, pCR scores were significantly greater in the ER-low group.

There are many studies in the literature that have investigated the relation among Ki67 levels and pCR. Rapoport *et al*. [27] showed in a research of 208 individuals that pCR rates were notably better in individuals with Ki67>40. A similar result was observed in a research of 522 patients by Fasching *et al*. [28]. The findings of the research by Akdag *et al*. [5] were consistent with those documented in the available research. Although the pCR rates of patients with Ki67 ≥ 20% were better in our study, they lost significance in multivariate analysis.

Tumour grade is another parameter for predicting pCR. Fisher *et al*. [29] showed that higher tumour grade was linked with greater pCR in a study of 1523 patients. Similar outcomes were seen in the Jones *et*

Table 4. Disease-free survival outcomes

		Whole cohort	Subgroup analysis		
			HER2-0	HERE2-Low	P value
Disease free survival	Median (months)	Not reached	Not Reached	Not reached	
	3 years (%)	85	88.7	83.2	0.12
	5 years (%)	74	86	69.6	
mFU (months)			37		

HER2= human epidermal growth factor receptor, FU=follow-up

al. [22]. Similarly, the study of Ring *et al*. [21] confirmed this. Although the pCR rates of patients with tumour grade 3 were better in our study, it lost its significance in multivariate analysis. The small number of patients may have reduced the statistical power of the study and may have caused this.

There are studies within the source material showing that individuals having HR (+) who had pCR got better survival outcomes than those who could not achieve pCR [13, 28]. A meta-analysis of 27895 patients by Spring *et al*. [30] also showed that the DFS was noticeably longer for those who attained pCR. In this current research, the median DFS of those who attained pCR was not reached, whereas the median DFS of those who attained pCR was 92 months, which was statistically significant.

Although there are mixed outcomes in the source material about the relationship among HER2-0 and HER2-low status and pCR, large meta-analyses suggest that the HER2-0 group has better pCR rates [20]. However, the similar pCR rates of both groups in our study and the existence of studies with similar results emphasise the requirement for more study to clarify the intricate relationship between HER2-low status and HER2-0 and other tumor features in determining prognosis and responsiveness to therapy. Based on the current clinical and pathological data, our study offers insightful information; nevertheless, more research will clarify the many biological traits and prognostic consequences linked to the HER2-0 and HER2-low subgroups.

Limitations

When evaluating the findings, it is important to take into account the many limitations of our study. First, our findings may not be as generalizable to broader populations due to its retrospective nature and single-center methodology. Second, the relatively small sample size is a limitation that may affect the statistical power of multivariate analyses and requires careful interpretation. Notwithstanding these drawbacks, we think that our research offers useful empirical information that will further knowledge of HER2-0 and HER2-low early BC.

CONCLUSION

Our study concludes that NACT-treated individuals who had early HR (+) HER2-low and HER2-0 BC had comparable DFS and pCR rates. In line with previous research, we found that those who attained pCR had longer DFS. More significantly, our multivariate analysis revealed that low ER % was linked to pCR on its own. To validate these findings, explore the underlying biological processes, and facilitate the creation of more individualized treatment plans for individuals in these categories, further extensive prospective investigations are required.

Ethics Approval and Consent to Participate

This study was approved by the Tokat Gaziosmanpasa University Faculty of Medicine Non-Interventional Scientific Research Ethics Committee (Decision no: 25-MOBAEK-043 and date: 04.02.2025), and it was realized in compliance with the Declaration of Helsinki's enets.

Data Availability

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

Authors' Contribution

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

Conflict of Interest

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

Financing

The authors disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgments

We would like to express our sincere gratitude to Prof. Dr. Banu Öztürk and Prof. Dr. Özge Gümüşay for their valuable contributions in establishing the oncology clinic at Tokat Gaziosmanpaşa University and for their dedicated care of the patients included in this study.

Generative Artificial Intelligence Statement

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

Editor's Note

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

REFERENCES

1. Siegel RL, Kratzer TB, Giaquinto AN, Sung H, Jemal A. Cancer statistics, 2025. CA Cancer J Clin. 2025;75(1):10-45. Epub 20250116. doi: 10.3322/caac.21871.
2. Trapani D, Ginsburg O, Fadelu T, et al. Global challenges and policy solutions in breast cancer control. Cancer Treat Rev. 2022;104:102339. doi: 10.1016/j.ctrv.2022.102339.
3. Schettini F, Martínez-Sáez O, Falato C, et al. Prognostic value of intrinsic subtypes in hormone-receptor-positive metastatic breast cancer: systematic review and meta-analysis. ESMO Open. 2023;8(3):101214. doi: 10.1016/j.esmoop.2023.101214.
4. Cejalvo JM, Martínez de Dueñas E, Galván P, et al. Intrinsic Subtypes and Gene Expression Profiles in Primary and Metastatic Breast Cancer. Cancer Res. 2017;77(9):2213-2221. doi: 10.1158/0008-5472.CAN-16-2717.
5. Akdag G, Yildirim S, Dogan A, et alNeoadjuvant Chemotherapy and Pathologic Complete Response in HR+/HER2- Breast Cancer: Impact of Tumor Ki67 and ER Status. Chemotherapy. 2024;69(3):141-149. doi: 10.1159/000537874.
6. Yau C, Osdoit M, van der Noordaa M, et al. Residual cancer burden after neoadjuvant chemotherapy and long-term survival outcomes in breast cancer: a multicentre pooled analysis of 5161 patients. Lancet Oncol. 2022;23(1):149-160. doi: 10.1016/s1470-2045(21)00589-1.
7. Boughey JC, Ballman KV, McCall LM, et al. Tumor Biology and Response to Chemotherapy Impact Breast Cancer-specific Survival in Node-positive Breast Cancer Patients Treated With Neoadjuvant Chemotherapy: Long-term Follow-up From ACOSOG Z1071 (Alliance). Ann Surg. 2017;266(4):667-676. doi: 10.1097/sla.0000000000002373.
8. Tarantino P, Viale G, Press MF, et al. ESMO expert consensus statements (ECS) on the definition, diagnosis, and management of HER2-low breast cancer. Ann Oncol. 2023;34(8):645-59. doi: 10.1016/j.annonc.2023.05.008.
9. Modi S, Jacot W, Yamashita T, et al.; DESTINY-Breast04 Trial Investigators. Trastuzumab Deruxtecan in Previously Treated HER2-Low Advanced Breast Cancer. N Engl J Med. 2022;387(1):9-20. doi: 10.1056/NEJMoa2203690.
10. Corti C, Giugliano F, Nicolò E, Tarantino P, Criscitiello C, Curigliano G. HER2-Low Breast Cancer: a New Subtype? Curr Treat Options Oncol. 2023;24(5):468-478. doi: 10.1007/s11864-023-01068-1.
11. Zhang H, Katerji H, Turner BM, Hicks DG. HER2-Low Breast Cancers. Am J Clin Pathol. 2022;157(3):328-336. doi: 10.1093/ajcp/aqab117.
12. Ogston KN, Miller ID, Payne S, et al. A new histological grading system to assess response of breast cancers to primary chemotherapy: prognostic significance and survival. Breast. 2003;12(5):320-327. doi: 10.1016/s0960-9776(03)00106-1.
13. Guarneri V, Broglio K, Kau SW, et al. Prognostic value of pathologic complete response after primary chemotherapy in relation to hormone receptor status and other factors. J Clin Oncol. 2006;24(7):1037-1044. doi: 10.1200/jco.2005.02.6914.
14. Zhou S, Liu T, Kuang X, et al. Comparison of clinicopathological characteristics and response to neoadjuvant chemotherapy between HER2-low and HER2-zero breast cancer. Breast. 2023;67:1-7. doi: 10.1016/j.breast.2022.12.006.
15. de Moura Leite L, Cesca MG, Tavares MC, et al. HER2-low status and response to neoadjuvant chemotherapy in HER2 negative early breast cancer. Breast Cancer Res Treat. 2021;190(1):155-163. doi: 10.1007/s10549-021-06365-7.
16. Denkert C, Seither F, Schneeweiss A, et al. Clinical and molecular characteristics of HER2-low-positive breast cancer: pooled analysis of individual patient data from four prospective, neoadjuvant clinical trials. Lancet Oncol. 2021;22(8):1151-1161. doi: 10.1016/S1470-2045(21)00301-6.
17. Baez-Navarro X, van Bockstal MR, Jager A, van Deurzen CHM. HER2-low breast cancer and response to neoadjuvant chemotherapy: a population-based cohort study. Pathology. 2024;56(3):334-342. doi: 10.1016/j.pathol.2023.10.022.
18. de Moraes FCA, de Castro Ribeiro CHD, Pessôa FDDL, et al. Pathologic response rates in HER2-low versus HER2-zero early breast cancer patients receiving neoadjuvant therapy: a systematic review and meta-analysis. Breast Cancer Res. 2025;27(1):39. doi: 10.1186/s13058-025-01989-9.
19. Yang M, Sun J, Liu L, et al. Clinicopathological characteristics of HER2-low breast cancer: a retrospective study. Sci Rep. 2023;13(1):12382. doi: 10.1038/s41598-023-39372-3.
20. Ilie SM, Briot N, Constantin G, et al. Pathologic complete response and survival in HER2-low and HER2-zero early breast cancer treated with neoadjuvant chemotherapy. Breast Cancer. 2023;30(6):997-1007. doi: 10.1007/s12282-023-01490-1.
21. Ring AE, Smith IE, Ashley S, Fulford LG, Lakhani SR. Oestrogen receptor status, pathological complete response and prognosis in patients receiving neoadjuvant chemotherapy for early breast cancer. Br J Cancer. 2004;91(12):2012-2017. doi: 10.1038/sj.bjc.6602235.
22. Jones RL, Salter J, A'Hern R, et al. Relationship between oestrogen receptor status and proliferation in predicting response and long-term outcome to neoadjuvant chemotherapy for breast cancer. Breast Cancer Res Treat. 2010;119(2):315-323. doi: 10.1007/s10549-009-0329-x.

23. Dieci MV, Griguolo G, Bottosso M, et al. Impact of estrogen receptor levels on outcome in non-metastatic triple negative breast cancer patients treated with neoadjuvant/adjuvant chemotherapy. NPJ Breast Cancer. 2021;7(1):101. doi: 10.1038/s41523-021-00308-7.
24. Gloyeske NC, Dabbs DJ, Bhargava R. Low ER+ breast cancer: Is this a distinct group? Am J Clin Pathol. 2014;141(5):697-701. doi: 10.1309/AJCP34CYSATWFDPQ.
25. Fujii T, Kogawa T, Dong W, et al. Revisiting the definition of estrogen receptor positivity in HER2-negative primary breast cancer. Ann Oncol. 2017;28(10):2420-2428. doi: 10.1093/annonc/mdx397.
26. Ohara AM, Naoi Y, Shimazu K, et al. PAM50 for prediction of response to neoadjuvant chemotherapy for ER-positive breast cancer. Breast Cancer Res Treat. 2019;173(3):533-543. doi: 10.1007/s10549-018-5020-7.
27. Rapoport BL, Barnard-Tidy J, Van Eeden RI, Smit T, Nayler S, Benn C. Pathological complete response in early breast cancer patients undergoing neoadjuvant chemotherapy: Focus on Ki-67 and molecular subtypes. Ann Oncol. 2019;30(Supplement 3):iii37. doi: 10.1093/annonc/mdz097.012.
28. Fasching PA, Heusinger K, Haeberle L, et al. Ki67, chemotherapy response, and prognosis in breast cancer patients receiving neoadjuvant treatment. BMC Cancer. 2011;11:486. doi: 10.1186/1471-2407-11-486.
29. Fisher ER, Wang J, Bryant J, Fisher B, Mamounas E, Wolmark N. Pathobiology of preoperative chemotherapy: findings from the National Surgical Adjuvant Breast and Bowel (NSABP) protocol B-18. Cancer. 2002;95(4):681-695. doi: 10.1002/cncr.10741.
30. Spring LM, Fell G, Arfe A, et al. Pathologic Complete Response after Neoadjuvant Chemotherapy and Impact on Breast Cancer Recurrence and Survival: A Comprehensive Meta-analysis. Clin Cancer Res. 2020;26(12):2838-2848. doi: 10.1158/1078-0432.CCR-19-3492.

The European Research Journal 2025;11(5):868-877 Original Article

DOI: https://doi.org/10.18621/eurj.1681181 *Obstetrics and Gynecology*

The impact of biochemical marker levels in pregnant patients diagnosed with HELLP syndrome on predicting the progression and recovery timelines

Mehmet Alican Sapmaz[1], Sait Erbey[1], Murat Polat[1], Ahmet Kurt[1], Dilara Sarıkaya Kurt[1], Kadriye Yakut Yücel[2], Recep Taha Ağaoğlu[2]

[1]Department of Obstetrics and Gynecology, Etlik City Hospital, Ankara, Türkiye; [2]Department of Perinatology, Etlik City Hospital, Ankara, Türkiye

ABSTRACT

Objectives: This study investigates the role of biochemical markers in predicting the clinical course of pregnant women with Hemolysis, Elevated Liver Enzymes and Low Platelets (HELLP) syndrome, aiming to correlate marker levels with disease progression and recovery timelines for improved prognostic assessment and therapeutic strategies.

Methods: A retrospective analysis was conducted on 50 pregnant patients (aged 18-45) diagnosed with HELLP syndrome between October 2022 and November 2023. Data on demographics, vital signs, clinical symptoms, and laboratory markers (platelets, liver enzymes, lactate dehydrogenase [LDH], and bilirubin) were examined. Outcomes measured included complications, intensive care unit needs, and recovery time.

Results: The mean age was 30.74±5.27 years and body mass index of 29.76±5.88 kg/m2. The gestational age was 31.97±4.45 weeks. Significant cut-off values were identified for urea at 27.50 (sensitivity: 100%, specificity: 73%, R2= 0.553, P<0.001) and creatinine at 0.85 (sensitivity: 100%, specificity: 91%, (P<0.001). LDH, bilirubin, and platelets also showed predictive value for clinical outcomes (P-values ranging from 0.005 to <0.05). Neutrophil-to-Lymphocyte Ratio and urea correlated with longer postpartum stays and complications, while higher mean platelet volume was linked to shorter stays (NLR: β=0.303, P=0.009; BUN: β=0.553, P<0.001).

Conclusions: The study highlights the importance of renal and hematological markers (urea, creatinine, LDH, bilirubin, platelets) in predicting HELLP outcomes. Renal markers showed high sensitivity, while hematological markers correlated with hospital stay duration, supporting their integration into clinical protocols to optimize treatment and patient management.

Keywords: Biomarkers, HELLP syndrome, hospital stay, predictive value

Hemolysis, Elevated Liver Enzymes and Low Platelets (HELLP) syndrome occurs in about 0.2% to 0.8% of pregnancies and is associated with increased risks of complications for both the mother and the fetus [1]. While hypertension is commonly observed, the symptoms of preeclampsia

Received: April 21, 2025 **Accepted:** August 21, 2025 **Available Online:** August 25, 2025 **Published:** September 4, 2025

How to cite this article: Sapmaz MA, Erbey S, Polat M, et al. The impact of biochemical marker levels in pregnant patients diagnosed with HELLP syndrome on predicting the progression and recovery timelines. Eur Res J. 2025;11(5):868-877. doi: 10.18621/eurj.1681181

Corresponding author: Mehmet Alican Sapmaz, MD., Phone: +90 312 797 00 00, E-mail: dr.alicansapmaz@hotmail.com

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

[2] can be either subtle or absent [3]. Prompt and precise diagnosis is crucial for effective management. Furthermore, the maternal symptoms can be non-specific and may be confused with other medical or obstetric issues, necessitating differential diagnosis [4]. HELLP syndrome is defined by two primary diagnostic frameworks. The more commonly used Tennessee classification defines the condition by the presence of microangiopathic hemolytic anemia, which is indicated by an abnormal blood smear, reduced serum haptoglobin, and increased levels of lactate dehydrogenase (LDH) (above 600 IU/L, more than twice the normal upper limit) and aspartate aminotransferase (AST) (above 70 IU/L). Additionally, bilirubin levels greater than 1.2 mg/dL and a platelet count below 100×10^9 are required for diagnosis [5]. A less severe variant, known as ''HELLP,'' meets only two of these criteria. Additionally, the Mississippi Triple-class classification system differentiates the syndrome based on the lowest platelet count observed [6].

HELLP syndrome is often associated with preeclampsia, a condition characterized by the onset of high blood pressure (≥140/90 mmHg after 20 weeks of gestation) that resolves after delivery, along with notable proteinuria (≥300 mg/day or a spot urine-to-creatinine ratio ≥30 mg/mmol) [2]. PE is about ten times more common than HELLP syndrome. When preeclampsia and HELLP occur before 28 weeks of gestation, which accounts for roughly 20-30% of cases, they are usually associated with more severe complications [7, 8]. The clinical presentation of HELLP can progress rapidly and may indicate severe complications [9].

The main goal of this study is to assess how biochemical markers can predict the clinical progression and recovery in pregnant patients diagnosed with HELLP syndrome. By examining the levels of these markers, the study seeks to identify correlations with the severity and trajectory of the syndrome.

METHODS

This retrospective cohort study sought to explore how the levels of biochemical markers in pregnant women diagnosed with HELLP syndrome influence the prediction of their progression and recovery timelines. Our study was commenced after the consent of Ankara Etlik City Hospital No. 1 Clinical Research Ethics Committee with the number: AEŞH-EK1-2023-728 on 06/12/2023. Our analysis was carried out following the ethical guidelines set forth in the Declaration of Helsinki. The study was conducted at the Department of Obstetrics and Gynecology and included patients treated between October 1st, 2022, and November 1st, 2023. The study cohort included all individuals diagnosed with HELLP syndrome within the designated study timeframe. To be included in the study, patients had to have a confirmed diagnosis of HELLP syndrome based on clinical and laboratory criteria and have completed treatment for the condition at our centre. Individuals were excluded from the study if they were younger than 18 years, had a history of systemic health conditions, were carrying multiple fetuses, had a previous diagnosis of cancer prior to pregnancy, were engaged in active smoking, alcohol consumption, or illicit drug use, were diagnosed with infectious diseases such as HIV, HCV, or HBV, had any chronic hepatic pathology, or had undergone liver transplantation in the past.

Data for the study was collected retrospectively from the medical records of eligible patients. This included demographic data such as maternal age, gravidity, parity, history of abortions, number of living children, and any documented chronic illnesses. Vital signs such as blood pressure, heart rate, respiratory rate, and body temperature were monitored during the patients' entire hospital admission. Detailed information regarding patients' clinical symptoms, including headache, epigastric pain, visual disturbances, nausea, and vomiting, was retrieved from their medical records. Comprehensive laboratory data was collected, including platelet count, hemoglobin levels, liver enzymes (Alanine aminotransferase [ALT] and AST), LDH levels, bilirubin levels, uric acid levels, creatinine (Cr) levels, and proteinuria levels. Documentation regarding the patients' clinical course, including duration of hospitalization, development of HELLP syndrome-related complications (e.g., disseminated intravascular coagulation (DIC), placental abruption, acute renal failure, eclampsia, cerebral hemorrhage, acute respiratory distress syndrome (ARDS), liver hematoma), need for intensive care unit admission, and requirement for blood product transfusions was meticulously reviewed and recorded. Postoperative laboratory values at 2 hours, 6 hours, and 24 hours

were also collected to assess the trajectory of biochemical markers.

The collected data was analyzed to investigate potential correlations between the initial levels of key biochemical markers (platelet count, ALT and AST, LDH, and bilirubin) and various aspects of HELLP syndrome. These aspects included the severity of the syndrome, the likelihood of developing complications, the overall recovery timeline, the duration of hospitalization, and the time required for blood pressure and other clinical parameters to normalize. This in-depth analysis sought to offer a broader insight into the prognostic significance of these biochemical markers in HELLP syndrome, with the potential to enhance clinical management of this severe pregnancy complication.

Statistical Analysis

Statistical analysis was conducted using the SPSS software (Version 27, SPSS Inc., Chicago, IL, USA). The normality of the distributions of biochemical markers (platelet count, ALT and AST, LDH, and bilirubin) was evaluated using the Kolmogorov–Smirnov test, Shapiro–Wilk test, and histograms. The relationship between the initial levels of these markers and the clinical parameters, including the duration of hospitalization, time to normalization of blood pressure, and the occurrence of HELLP syndrome-related complications was assessed. Spearman's correlation analysis was used to assess the association between the initial levels of biochemical markers and the severity of HELLP syndrome, as reflected by factors such as hospitalization duration, time to blood pressure normalization, and the occurrence of complications. A P-value of less than 0.05 was considered statistically significant for all analyses.

RESULTS

In our study, demographic and clinical characteristics of patients diagnosed with HELLP syndrome were presented (Table 1). Hematologic and biochemical parameters in patients with HELLP syndrome are summarized in Table 1. Patients demonstrated significant abnormalities in several markers. The mean systolic blood pressure was elevated at 164.8±18.1 mmHg, and the mean diastolic blood pressure was 104.1±12.5 mmHg, indicating prominent hypertension. Lactate dehydrogenase (LDH) levels were markedly increased, with a mean value of 730.3±589.3 U/L, consistent with hemolysis. Total bilirubin averaged 1.1±1.3 mg/dL, and direct bilirubin was elevated at 0.6±0.8 mg/dL, reflecting liver dysfunction. Other notable findings included a mean neutrophil count of 10.883±3,743/mm³, a neutrophil-to-lymphocyte ratio (NLR) of 6.75±4.73, and reduced platelet counts averaging 137.3±92.4 ×10³/mm³. Elevated liver enzymes were also observed, with alanine aminotransferase (ALT) at 177.4±225.0 IU/L and aspartate aminotransferase (AST) at 276.0±367.7 IU/L.

The white blood cell count was notably elevated, especially in neutrophils, which contributed to a high neutrophil-to-lymphocyte ratio (NLR), a key indicator of systemic inflammation. Platelet counts were significantly reduced, a hallmark of thrombocytopenia in HELLP syndrome. Additionally, elevated ALT and AST further supported the diagnosis of hepatic involvement. Renal function was also impacted, as evidenced by elevated blood urea nitrogen (BUN) and creatinine levels, suggesting some degree of renal compromise. These findings reflect the complex nature of HELLP syndrome, involving significant hemolysis, liver dysfunction, and renal involvement.

Maternal and neonatal outcomes in patients diagnosed with HELLP syndrome were thoroughly evaluated (Table 2). Eclampsia occurred in 16% of cases, representing a significant complication. Thrombocytopenia was observed in 8% of patients with varying severity, while the majority (92%) had platelet counts that did not require extensive transfusion. Fibrinogen levels remained within normal ranges in 96% of patients. Maternal admission to the intensive care unit was required in 16% of cases, though no maternal deaths were reported. Postpartum hemorrhage was noted in 6% of patients.

Neonatal outcomes in HELLP syndrome revealed significant complications (Table 2). Intrauterine growth restriction (IUGR) was observed in 38.8% of pregnancies, while preterm birth or preterm premature rupture of membranes (PPROM) occurred in 82% of cases. External cephalic version was frequently performed (data not explicitly provided but implied by interventions). Fetal distress was noted in 16% of pregnancies. Neonatal intensive care unit (NICU) admission was required for 83.7% of neonates, with 52%

Table 1. Demographic, clinical characteristics, hematological, and biochemical parameters of patients with HELLP syndrome

Characteristics	Data
Gestational age (weeks)	31.9±4.5
Age (years)	30.7±5.3
BMI (kg/m^2))	29.8±5.9
Parity	
0	24 (48%)
1	18 (36%)
2	4 (8%)
3	1 (2%)
4	2 (4%)
7	1 (2%)
Smoke	
No	48 (96%)
Yes	2 (4%)
IVF	
No	47 (94%)
Yes	3 (6%)
GDM	
No	45 (91.8%)
Yes	4 (8.2%)
Delivery	
Normal spontaneous delivery	3 (6%)
Cesarean section	47 (9%4)
Systolic blood pressure (mmHg)	164.8±18.1
Diastolic blood bressure (mmHg)	104.1±12.5
LDH (U/L)	730.3±589.29
Total bilirubin (mg/dL)	1.1±1.3
Direct bilirubin (mg/dL)	0.6±0.8
White blood cell (/mm^3)	14035±4309.1
Neutrophil (/mm^3)	10883.2±3742.99
Lymphocyte (/mm^3)	2261.8±1530.44
NLR	6.75±4.73
Monocyte (/mm^3)	762.4±289.6
Haematocrit (%)	37.65±5.97
MPV (fL)	11.21±1.07
Platelets (×1000/mm^3)	137.32±92.37
PLR	0.08±0.07
Haemoglobin (g/dL)	12.39±1.9
ALT (IU/L)	177.38±225.01
AST (IU/L)	275.98±367.66
BUN (mg/dL)	29.21±19.83
Creatinine (mg/dL)	1.12±2.5

Data are shown as mean±standard deviation or n (%). BMI=Body Mass Index, IVF=In-vitro Fertilisation, GDM=Gestational Diabetes Mellitus, HELLP=Hemolysis, Elevated Liver Enzymes and Low Platelets, LDH=lactate dehydrogenase, NLR=Neutrophil-to-Lymphocyte Ratio, MPV=Mean Platelet Volume, PLR=Platelet-to-lymphocyte ratio, ALT=Alanine Aminotransferase, AST=Aspartate aminotransferase, BUN=Blood Urea Nitrogen

Table 2. Maternal and neonatal outcomes in HELLP syndrome

	Frequency of Occurrence / Application	Data
Erythrocyte suspension fransfusion	0	43 (86%)
	2	4 (8%)
	3	1 (2%)
	4	2 (4%)
Transfusion of FFP	0	45 (90%)
	2	4 (8%)
	20	1 (2%)
Platelet (×1000/mm^3)	0	46 (92%)
	1	1 (2%)
	2	2 (4%)
	3	1 (2%)
Fibrinogen	0	48 (96%)
	1	2 (4%)
Maternal intensive care unit)	0	42 (84%)
	1	8 (16%)
Maternal death	0	50 (100%)
Postpartum haemorrhage	0	47 (94%)
	1	3 (6%)
IUGR	0	30 (61.2%)
	1	19 (38.8%)
Preterm/ PPROM	0	9 (18%)
	1	41 (82%)
Intrauterine ex fetus	0	49 (98%)
	1	1 (2%)
Fetal distress	0	42 (84%)
	1	8 (16%)
Neonatal intensive care unit	0	8 (16.3%)
	1	41 (83.7%)
Respiratory distress syndrome	0	24 (48%)
	1	26 (52%)
Neonatal sepsis	0	43 (86%)
	1	7 (14%)
Neonatal pneumonia	0	44 (88%)
	1	6 (12%)
Retinal hemorrhage/ ROP	0	39 (78%)
	1	11 (22%)
Intraventricular hemorrhage	0	47 (94%)
	1	3 (%6)
Hypoxic-ischemic encephalopathy	0	50 (100%)
Necrotizing enterocolitis	0	46 (92%)
	1	3 (6%)
	10	1 (2%)
Early neonatal death	0	45 (90%)
	1	5 (10%)

Data are shown as n (%). FFP=fresh frozen plasma, IUGR=intrauterine growth retardation, HELLP=Hemolysis, Elevated Liver Enzymes and Low Platelets, PPROM=preterm premature rupture of membranes, ROP=retinopathy of prematurity

Table 3. Hospital stay duration and neonatal outcomes

	Data
Before birth hospital stay (hours)	29.12±67.11
After birth hospital stay (hours)	147.54±91.37
Birth weight (gram)	1705.3±876.02
APGAR 1	5.92±2.51
APGAR 5	7.8±2.06
Hospitalization (day)	22.98±26.63

Data are shown as mean±standard deviation.

experiencing respiratory distress syndrome (RDS). Neonatal sepsis and pneumonia affected 14% and 12% of neonates, respectively. Additionally, retinopathy of prematurity (ROP) occurred in 22%, and intraventricular hemorrhage (IVH) in 6% of neonates. Hypoxic-ischemic encephalopathy (HIE) was universal among the cohort, with 100% requiring intervention, often including therapeutic hypothermia. Necrotizing enterocolitis affected 8% of neonates, while early neonatal death occurred in 10%, underscoring the severe neonatal morbidity and mortality associated with HELLP syndrome.

Data on the duration of hospitalization and neonatal outcomes related to HELLP syndrome are presented in Table 3. The average duration of hospitalization before birth was relatively short, measured at 29.12±67.11 hours, while the post-birth hospital stay was substantially longer, averaging 147.54±91.37 hours. Neonatal outcomes revealed a high prevalence of low-birth-weight infants, with a mean birth weight of 1705.3±876.02 grams. The mean APGAR score at 1 minute was 5.92±2.51, indicating that many infants required initial resuscitative support; however, the 5-minute APGAR score improved significantly to 7.8±2.06, suggesting stabilization shortly after birth. Neonatal hospitalization duration averaged 22.98±26.63 days, with variability suggesting prolonged NICU stays in some cases. These findings underscore the critical importance of intensive neonatal care and close monitoring in pregnancies affected by

Table 4. Diagnostic accuracy of biomarkers in eclampsia* and related complications in posterior reversible encephalopathy syndrome (PRES)

Eclampsia	AUC	SD	P value	95% CI		Cut-off	Sensitivity	Specificity
				Lower	Upper			
Total bilirubin	0.728*	0.079*	**0.022***	0.573*	0.884*	0.49*	0.82*	0.69*
	0.689	0.090	**0.044**	0.513	0.866	0.49	0.77	0.70
Direct bilirubin	0.760*	0.075*	**0.009***	0.613*	0.907*	0.19*	0.73*	0.77*
	0.715	0.088	**0.022**	0.543	0.888	0.19	0.69	0.78
WBC	0.702*	0.106*	**0.043***	0.493*	0.910*	16120*	0.64*	0.82*
	0.713	0.092	**0.023**	0.533	0.893	14755	0.69	0.73
Lymphocyte	0.797*	0.084*	**0.003***	0.633*	0.961*	2245*	0.73*	0.77*
	0.686	0.100	**0.048**	0.489	0.883	2245	0.62	0.76
NLR	0.732*	0.082*	**0.020***	0.571*	0.893*	4.67*	0.64*	0.72*
Platelets	0.852*	0.054*	**<0.001***	0.746*	0.958*	165.5*	0.82*	0.79*
	0.772	0.081	**0.004**	0.614	0.931	134.5	0.85	0.73
ALT	0.825*	0.087*	**0.001***	0.655*	0.995*	73.5*	0.82*	0.79*
	0.794	0.096	**0.002**	0.605	0.983	73.5	0.77	0.81
AST	0.791*	0.087*	**0.003***	0.621*	0.962*	68*	0.73*	0.85*
	0.756	0.098	**0.007**	0.564	0.947	68	0.69	0.86

WBC=white blood cell, NLR=Neutrophil-to-Lymphocyte Ratio, ALT=Alanine aminotransferase, AST=aspartate aminotransferase, AUC=Area Under Curve, SD=Standart Deviation, CI= Confidence Interval, *Statistically significant difference or result

Table 5. Biomarkers for hepatic ischemia and rupture

Hepatic ischemia-rupture	AUC	SD	P value	95% CI		Cut-off	Sensitivity	Specificity
				Lower	Upper			
LDH	0.710	0.078	**0.011**	0.558	0.863	572	0.74	0.74
Total bilirubin	0.729	0.075	**0.006**	0.582	0.875	0.59	0.78	0.74
Direct bilirubin	0.733	0.072	**0.005**	0.592	0.875	0.28	0.70	0.70
Lymphocyte	0.709	0.074	**0.011**	0.564	0.855	1765	0.74	0.70
PLatelets	0.665	0.078	**0.046**	0.512	0.818	94	0.70	0.63
ALT	0.870	0.055	**<0.001**	0.762	0.977	100	0.87	0.85
AST	0.845	0.059	**<0.001**	0.729	0.962	146	0.87	0.81

LDH=Lactate dehydrogenase, ALT=Alanine aminotransferase, AST=Aspartate aminotransferase, AUC=Area Under Curve, SD=Standart Deviation, CI= Confidence Interval

HELLP syndrome.

The diagnostic accuracy of various biomarkers for eclampsia and related complications, including Posterior Reversible Encephalopathy Syndrome (PRES), was assessed. Among the biomarkers evaluated, platelet count exhibited the highest discriminatory power, with an AUC of 0.852 (P<0.001), sensitivity of 82%, and specificity of 79% at a cut-off value of 165.5 $\times 10^3/mm^3$, suggesting strong clinical relevance as a diagnostic tool. ALT also demonstrated excellent predictive value with an AUC of 0.825 (P=0.001), sensitivity of 82%, and specificity of 79% at a cut-off of 73.5 U/L. Both platelet count and ALT were identified as reliable and statistically significant indicators for the diagnosis of eclampsia, with high sensitivity and specificity values that support their integration into clinical risk assessment models (Table 4).

In women with PRES, ALT and AST demonstrated strong predictive abilities for cerebral edema and stroke related to eclampsia. ALT was a strong predictor, with a high sensitivity and specificity, while AST also showed notable diagnostic performance. Platelet, while moderately predictive, still proved useful in identifying eclampsia-related complications, with a reasonable balance between sensitivity and specificity. These findings suggest that a combination of specific biomarkers - particularly platelets (AUC= 0.852, P<0.001), ALT (AUC 0.825, P=0.001), AST (AUC=0.791, P=0.003), and lymphocyte count (AUC=0.797, P=0.003) - may serve as highly valuable clinical tools for the diagnosis and management of eclampsia and its associated complications, including PRES. These markers demonstrated strong diagnostic performance with high sensitivity (up to 85%) and specificity (up to 86%) (Table 4).

The evaluation of biomarkers for hepatic ischemia and rupture revealed significant findings. ALT demonstrated the strongest predictive value, with high sensitivity and specificity, making it a reliable indicator for diagnosing hepatic ischemia and rupture. AST also exhibited strong predictive performance, although slightly lower than ALT, providing valuable diagnostic insights. LDH, total bilirubin, and direct bilirubin showed moderate predictive value, highlighting their potential utility in the diagnosis of hepatic complications. Lymphocyte count and platelet count were less

Table 6. Association of biomarkers with hospital stays after birth

		β	t	P value	R^2
Hospital stays after birth	NLR	0.303	2.730	**0.009**	0.427
	BUN	0.553	4.987	**<0.001**	

NLR=Neutrophil-to-Lymphocyte Ratio, BUN=Blood Urea Nitrogen

effective in predicting hepatic ischemia and rupture, demonstrating lower sensitivity and specificity. These findings suggest that ALT and AST are particularly valuable biomarkers for predicting hepatic ischemia and rupture, demonstrating the highest diagnostic performance with AUC values of 0.870 (P<0.001) and 0.845 (P<0.001), respectively. In contrast, biomarkers such as LDH (AUC= 0.710, P=0.011), total bilirubin (AUC= 0.729, P=0.006), direct bilirubin (AUC= 0.733, P=0.005), and lymphocyte count (AUC=0.709, P=0.011) showed moderate predictive value and may serve as supplementary diagnostic tools (Table 5).

The analysis of factors influencing hospital stays after birth revealed significant correlations with various biomarkers. The NLR showed a moderate positive correlation with the duration of hospital stays, indicating that higher NLR values are associated with longer stays. However, the relatively modest correlation suggests that NLR alone accounts for only a limited portion of the variability in hospital stay duration. In contrast, BUN demonstrated a stronger positive correlation with the length of hospital stay, suggesting that elevated BUN levels are a more substantial predictor of longer hospitalizations. These findings highlight BUN as a significantly more reliable marker than NLR for predicting the length of post-birth hospital stay, with a higher standardized beta coefficient (β= 0.553, P<0.001) compared to NLR (β=0.303, P=0.009), as shown in Table 6. The model explained 42.7% of the variance in hospital stay duration (R^2=0.427).

DISCUSSION

Our most significant finding was that elevated levels of liver enzymes—particularly ALT and AST—demonstrated strong predictive value not only for the development of HELLP syndrome but also for specific complications such as eclampsia, cerebral edema, hepatic ischemia, and extended postpartum hospitalization. HELLP syndrome, a serious pregnancy disorder marked by hemolysis, increased liver enzyme levels, and thrombocytopenia, presents considerable risks to both the mother and the fetus. It is often associated with adverse outcomes such as eclampsia, placental abruption, acute kidney injury, pulmonary edema, and maternal and fetal death [10, 11]. While the exact pathophysiology of HELLP syndrome remains unclear, systemic inflammation and endothelial dysfunction are believed to play key roles [12]. Predicting the progression of HELLP syndrome and its associated complications is crucial for timely intervention and improved patient management. This study is the first to thoroughly assess the predictive significance of various biochemical markers for specific complications of HELLP syndrome, such as eclampsia, cerebral edema and stroke, hepatic ischemia, and the length of postpartum hospital stay.

Prior research has identified that factors such as genetics, immune function, inflammation, metabolism, and coagulation processes play integral roles in the development of HELLP syndrome [13].The pathogenesis of this condition involves a complex interplay of factors, where inflammatory processes and immune reactions are critical. In these reactions, cells such as neutrophils, lymphocytes, and platelets are actively involved, secreting inflammatory cytokines that contribute to the syndrome's onset [13]. This study corroborates our findings regarding the association between elevated neutrophil and monocyte counts and HELLP syndrome. Both studies observed significantly higher levels of these cells in HELLP patients at the time of delivery compared to low-risk pregnancies. However, while this study focuses on the predictive value of various inflammatory indices derived from these cell counts in the first trimester, our research highlights the significance of liver enzymes, particularly ALT and AST, not just for predicting HELLP development but also for prognosticating specific complications [14]. Our findings suggest that routine monitoring of ALT and AST throughout pregnancy could offer valuable insights into HELLP progression, potentially beyond the predictive capacity of early pregnancy inflammatory indices. This difference in focus stems from our study's broader exploration of biochemical markers and their association with specific HELLP-related complications, offering a more comprehensive understanding of the syndrome's prediction and prognosis.

Sisti *et al*. [15] also investigated the role of inflammatory markers in HELLP syndrome, focusing on NLR and Platelet-to-lymphocyte ratio (PLR). Like our findings, they observed a significantly higher NLR and lower PLR in women with HELLP compared to healthy controls. However, their study primarily fo-

cused on these ratios as potential diagnostic markers for HELLP syndrome, while our research delved deeper into the predictive capacity of a broader range of biochemical markers, including liver enzymes, for specific HELLP-related complications. By evaluating outcomes like eclampsia, cerebral edema, hepatic ischemia, and postpartum hospital stay duration, our research offers a more extensive understanding of the prognostic value of these markers beyond their diagnostic potential.

Strengths and Limitations

This research offers important perspectives on the predictive significance of different biochemical markers for complications associated with HELLP syndrome. Nevertheless, it is crucial to recognize the study's limitations. Being a retrospective analysis, it is prone to inherent biases linked to data gathering and patient selection. The limited sample size may also restrict the ability to generalize the findings. Furthermore, the study design does not allow for definitive conclusions regarding causality between the observed biomarkers and specific HELLP syndrome complications.

Despite these limitations, our study has several strengths. It is the first to comprehensively evaluate a wide range of biochemical markers in relation to specific HELLP syndrome complications, providing a more nuanced understanding of their prognostic value. The study also highlights the importance of monitoring liver enzymes throughout pregnancy, potentially offering a valuable tool for identifying patients at higher risk for specific complications. These findings warrant further investigation in larger prospective studies to confirm the identified associations and elucidate the underlying mechanisms.

CONCLUSION

This study highlights the significant predictive value of biochemical markers in HELLP syndrome, particularly ALT and AST, renal markers (urea and creatinine), and inflammatory indices (NLR). Routine monitoring of these markers can offer valuable insights into disease progression and guide clinical management, ultimately improving maternal and fetal outcomes. Additional studies with larger sample sizes and prospective designs are necessary to confirm these results and investigate the underlying mechanisms of HELLP syndrome.

Ethics Approval and Consent to Participate

This study was approved by the Ankara Etlik City Hospital No. 1 Clinical Research Ethics Committee (Decision No: AEŞH-EK1-2023-728; date: 06.12.2023). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Informed consent was waived because of the retrospective nature of the study and the analysis used anonymous clinical data.

Data Availability

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

Authors' Contribution

Study Conception: MAS, AK; Study Design: MP, SE; Supervision: kky; Funding: RTA, AK; Materials: N/A; Data Collection and/or Processing: RTA, AK; Statistical Analysis and/or Data Interpretation: RTA, AK; Literature Review: DSK; Manuscript Preparation: MAS; and Critical Review: MP, SE.

Conflict of Interest

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

Financing

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

Acknowledgments

The authors have no acknowledgments to declare.

Generative Artificial Intelligence Statement

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

Editor's Note

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

REFERENCES

1. Adorno M, Maher-Griffiths C, Grush Abadie HR. HELLP Syndrome. Crit Care Nurs Clin North Am. 2022;34(3):277-288. doi: 10.1016/j.cnc.2022.04.009.
2. Lazo-Vega L, Toledo-Jaldin L, Badner A, et al. ACOG and local diagnostic criteria for hypertensive disorders of pregnancy (HDP) in La Paz-El Alto, Bolivia: A retrospective case-control study. Lancet Reg Health Am. 2022;9:100194. doi: 10.1016/j.lana.2022.100194.
3. Ives CW, Sinkey R, Rajapreyar I, Tita ATN, Oparil S. Preeclampsia-Pathophysiology and Clinical Presentations: JACC State-of-the-Art Review. J Am Coll Cardiol. 2020;76(14):1690-1702. doi: 10.1016/j.jacc.2020.08.014.
4. Turbeville HR, Sasser JM. Preeclampsia beyond pregnancy: long-term consequences for mother and child. Am J Physiol Renal Physiol. 2020;318(6):F1315-F1326. doi: 10.1152/ajprenal.00071.2020.
5. Ditisheim A, Sibai BM. Diagnosis and Management of HELLP Syndrome Complicated by Liver Hematoma. Clin Obstet Gynecol. 2017;60(1):190-197. doi: 10.1097/GRF.0000000000000253.
6. Haram K, Svendsen E, Abildgaard U. The HELLP syndrome: clinical issues and management. A Review. BMC Pregnancy Childbirth. 2009;9:8. doi: 10.1186/1471-2393-9-8.
7. James JL, Whitley GS, Cartwright JE. Pre-eclampsia: fitting together the placental, immune and cardiovascular pieces. J Pathol. 2010;221(4):363-378. doi: 10.1002/path.2719.
8. Habli M, Eftekhari N, Wiebracht E, Bombrys A, Khabbaz M, How H, Sibai B. Long-term maternal and subsequent pregnancy outcomes 5 years after hemolysis, elevated liver enzymes, and low platelets (HELLP) syndrome. Am J Obstet Gynecol. 2009;201(4):385.e1-5. doi: 10.1016/j.ajog.2009.06.033.
9. Lisonkova S, Razaz N, Sabr Y, et al. Maternal risk factors and adverse birth outcomes associated with HELLP syndrome: a population-based study. BJOG. 2020;127(10):1189-1198. doi: 10.1111/1471-0528.16225.
10. Sibai BM. Diagnosis, controls, and management of the syndrome of hemolysis, elevated liver enzymes, and low platelet count. Obstet Gynecol. 2004;103(5 Pt 1):981-991. doi: 10.1097/01.AOG.0000126245.35811.2a.
11. Abildgaard U, Heimdal K. Pathogenesis of the syndrome of hemolysis, elevated liver enzymes, and low platelet count (HELLP): a review. Eur J Obstet Gynecol Reprod Biol. 2013;166(2):117-123. doi: 10.1016/j.ejogrb.2012.09.026.
12. Young BC, Levine RJ, Karumanchi SA. Pathogenesis of preeclampsia. Annu Rev Pathol. 2010;5:173-192. doi: 10.1146/annurev-pathol-121808-102149.
13. Landi B, Tranquilli AL. HELLP syndrome and placental inflammatory pathology. Minerva Ginecol. 2008;60(5):389-398.
14. İpek G, Tanaçan A, Ağaoğlu Z, Peker A, Şahin D. Can SIRI or other inflammatory indices predict HELLP syndrome in the first trimester? J Reprod Immunol. 2023;159:104126. doi: 10.1016/j.jri.2023.104126.
15. Sisti G, Faraci A, Silva J, Upadhyay R. Neutrophil-to-Lymphocyte Ratio, Platelet-to-Lymphocyte Ratio, and Routine Complete Blood Count Components in HELLP Syndrome: A Matched Case Control Study. Medicina (Kaunas). 2019;55(5):123. doi: 10.3390/medicina55050123.

The European Research Journal 2025;11(5):878-884 Original Article

DOI: https://doi.org/10.18621/eurj.1729441 *General Surgery*

Evaluation of postoperative chronic neuropathic pain in patients with Lichtenstein hernia repair and laparoscopic hernia repair

Mehmet Fatih Erol[1], Muhammed Doğangün[2], Berkay Demir[3], Sinan Arıcı[4], Füsun Gözen[5]

[1]Department of General Surgery, Bursa Yüksek İhtisas Training and Research Hospital, Bursa, Türkiye; [2]Department of General Surgery, Sakarya Sadıka Sabancı State Hospital, Sakarya, Türkiye; [3]Department of Gastrointestinal Surgery, Ankara Bilkent City Hospital, Ankara, Türkiye; [4]Department of General Surgery, Bursa VM Medical Park Hospital, Bursa, Türkiye; [5]Department of Anesthesia and Reanimation, Bursa Yüksek İhtisas Research and Training Hospital, Bursa, Türkiye

ABSTRACT

Objectives: The high incidence of inguinal hernia over the globe makes hernia repair the most common procedure in general surgery, and 10-15% of all surgeries are composed of hernia repair. In this study, we tried to elucidate the effect of risk factors of laparoscopic techniques on chronic neuropathic pain formation in the postoperative period in patients who underwent Lichtenstein hernia repair and laparoscopic hernia repair.

Methods: A total of 404 patients have been enrolled in this study. Two different surgery techniques have been conducted on the participants, as Lichtenstein repair (n=214) and laparoscopic inguinal hernia repair (n=190). Demographic data of the patients were recorded. Transabdominal preperitoneal (TAPP) and totally extraperitoneal (TEP) techniques were applied to patients who underwent laparoscopic repair. The 'Lanss Pain Score' was utilized in the evaluation of chronic pain in the postoperative period.

Results: We detected a statistically significant difference between the groups in terms of age and the Leeds Assessment of Neuropathic Symptoms and Signs (LANSS) scores ($P<0.05$). It was determined that the mean age of the patients with Lichtenstein repair was 57, and the mean LANSS score was 3. It was determined that the mean age of the patients with laparoscopic inguinal hernia repair was 49, and the mean LANSS score was 1. The distribution of demographic and clinical findings of the patients who underwent laparoscopic inguinal hernia repair according to the operation methods revealed no statistically significant difference in patients who underwent TAPP and TEP methods.

Conclusions: In conclusion, the TEP and TAPP methods have lower pain in the postoperative period.

Keywords: Lichtenstein repair, laparoscopic inguinal hernia repair, transabdominal preperitoneal, totally extraperitoneal

It is thought that more than 20 million inguinal hernia repairs are performed annually worldwide [1]. Bassini made the first modern surgical treatment description for inguinal hernias in 1884, and over time, it became a technique created via using the patient's own tissue and successfully applied for many decades [2].

Since then different techniques have been intro-

Received: June 28, 2025 **Accepted:** August 18, 2025 **Available Online:** August 22, 2025 **Published:** September 4, 2025

How to cite this article: Erol MF, Doğangün M, Demir B, Arıcı S, Gözen F. Evaluation of postoperative chronic neuropathic pain in patients with Lichtenstein hernia repair and laparoscopic hernia repair. Eur Res J. 2025;11(5):878-884. doi: 10.18621/eurj.1729441

Corresponding author: Mehmet Fatih Erol, MD., Phone: +90 224 295 50 00, E-mail: mehmetfatiherol@yahoo.com

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

duced, and today laparoscopic methods have been widely used in inguinal hernia repairs. The most important advantages of laparoscopic inguinal hernia repairs are lower postoperative pain and reduced infection risk compared to open surgery, patients returning to their activities earlier, better cosmetic appearance, and similar results in terms of recurrence compared to open surgery. Today, the most commonly used laparoscopic methods in inguinal hernia repairs are: totally extraperitoneal (TEP) and transabdominal preperitoneal (TAPP) approaches [3, 4].

After the 1970s, the use of synthetic patches in inguinal hernia repairs to reduce tension and recurrence had a very important place in the development of modern hernia surgery. In the following years, many different methods have been used in inguinal hernia repairs [5]. The first laparoscopic inguinal hernia repair was performed by Ger in 1982 after detecting an inguinal hernia in a patient who had been operated for other reasons [6]. Standard inguinal hernia repair changed little over the hundred years before the use of synthetic mesh. The next major change was the initiation of laparoscopic repair. There are discussions in the literature about which approach should be performed for routine inguinal hernia repair [5].

Although various techniques have been reported for hernia repair, the "tension-free" hernia repair described by Lichtenstein is the preferred method among open surgery techniques [7]. With the application of laparoscopic surgery in inguinal hernia repairs, hernia surgery has gained a different dimension and these laparoscopic methods have become accepted all over the world and successfully applied in many centers in a short time. TEP and TAPP are tension-free methods, also has the general advantages of laparoscopic surgery such as less postoperative pain, shorter recovery time and good cosmetic result [3, 5].

Chronic pain is a potential complication following inguinal hernia repair, defined as pain lasting between three and six months with a fluctuating nature. Chronic neuropathic pain, also referred to as persistent pain, is characterized by symptoms such as burning or shooting sensations and may result from damage to the somatosensory system [8]. Both modern open and laparoscopic repairs of groin hernias have been associated with entrapment-related symptoms involving the genital nerves, particularly the genitofemoral and ilioinguinal nerves.

This type of pain negatively impacts physical activity and overall comfort, leading to a reduced quality of life [9]. While the exact incidence remains debated, evidence consistently shows that chronic neuropathic pain is becoming more prevalent than hernia recurrence, historically the most common complication of hernia repair [10].

Chronic neuropathic pain is typically sharp, activity-related, and localized to the groin, though it may radiate toward the inner thigh. Associated symptoms include paresthesia, hypoesthesia, and hyperesthesia [11]. The condition can arise from intraoperative or postoperative injury to the inguinal nerves. Intraoperative injuries may result from thermal damage or nerve compression caused by suture or tack fixation. Postoperative nerve damage may occur due to scar tissue formation or involvement with meshoma [11, 12]. This highlights the importance of meticulous surgical techniques and careful postoperative monitoring to minimize nerve damage and manage pain effectively.

The Leeds Assessment of Neuropathic Symptoms and Signs (LANSS) score is one of the neuropathic pain rating scales. In previous studies, we observed that LANSS has been applied to evaluate cervical hernia, small bowel obstruction, and different surgeries, but it has not been studied before for inguinal hernia surgery to our knowledge [13, 14].

Success of inguinal hernia repair is evaluated according to recurrence rates and the presence of chronic pain. In previous literature it has been reported that TEP and TAPP methods has less recurrence and less chronic pain compared to other methods [3, 15, 16]. In this study, we tried to elucidate the effect of risk factors of inguinal hernia repair techniques on chronic neuropathic pain formation in the postoperative period in patients who underwent Lichtenstein hernia repair and laparoscopic hernia repair.

METHODS

This study evaluated the medical records of 482 patients who underwent inguinal hernia repair between January 2018 and December 2020. All participants were at least six months postoperative at the time of data collection. Two different surgery techniques have been conducted on the participants as Lichtenstein repair (n=214) and laparoscopic inguinal hernia repair

(n=190). Patients were excluded if they had undergone additional lower abdominal or perineal surgery for other indications or if they did not attend regular postoperative check-ups. Additionally, patients with femoral hernia, previous inguinal hernia surgery, and urgent hernia surgery due to peritoneal sign or incarceration were excluded from the study (totally 78 patients).

The ethics committee approval has been granted at 29.12.2021 and protocol number 2011-KAEK-25 2021/12-22. The study has been performed according to the Declaration of Helsinki as a statement of ethical principles for medical research involving human subjects and informed consent has been obtained from all individuals.

The patients who participated in the study were operated by the same surgical team. Demographic data of the patients were recorded. TAPP and TEP techniques were applied to patients who underwent laparoscopic repair. Prolene mesh has been used in both Lichtenstein repair and laparoscopic repair. In the Lichtenstein repair, the mesh was fixed with sutures. In patients who underwent laparoscopic repair, the mesh was fixed to the pubis with a laparoscopic vicryl stapler. The 'LANSS Pain Score' was utilized in the evaluation of chronic pain in the post-operative 6th-month follow-up.

The Self-administered-LANSS (S-LANSS) score is a self-report version of the LANSS. The S-LANSS aims to identify pain of predominantly neuropathic origin, as distinct from nociceptive pain, without the need for clinical examination. The LANSS score is calculated between 0–24, <12 is considered no chronic pain, and a score of ≥12 is considered chronic pain.

Lichtenstein hernia repair was performed with an open technique and tension-free repair with an anterior approach to the anatomical area of the hernia. Prolene mesh was used in the repair. The mesh is fixed primarily with prolene sutures.

Laparoscopic hernia repair includes two different techniques, TEP and TAPP. In both techniques, the hernia area is reached with a posterior approach. Again, in both techniques, one 10mm camera trocar and two 5mm trocars were used. In TEP, trocars are placed in the preperitoneal space. In TAPP, the trocars are in the abdomen. Prolene mesh was used in the repair. Prolene mesh is fixed to the pubis with only one vicryl tacker.

Statistical Analysis

Patient data collected within the scope of the study were analyzed with the IBM Statistical Package for the Social Sciences (SPSS) for Windows 23.0 package program. Frequency and percentage for categorical data, mean and standard deviation for continuous data were given as descriptive values. "Independent Sample T-test" was utilized for comparisons between groups, and "Fisher's Exact or Pearson Chi-Square Test" was used for comparison of categorical variables. The results were considered statistically significant when the P value was less than 0.05.

RESULTS

The study population was divided into two different study groups according to the operation method. The Lichtenstein repair group consisted of 214 patients with a median age of 57±16 years. The majority of the individuals were male, 89.3% (n=191) and 10.7% (n=23) of them were female. In terms of gender dis-

Table 1. Distribution of demographic and clinical findings by operation types

Characteristics (n=404)	Lichtenstein repair (n=214)	Laparoscopic inguinal hernia repair (n=190)	P value
Age (years)	51±16	49±14	0.254
LANSS score	3±4	1±3	**<0.001**
Gender			0.553
Male	191 (89.3%)	165 (86.8%)	
Female	23 (10.7%)	25 (13.2%)	

Data are shown as mean±standard deviation or n (%). LANSS=Leeds Assessment of Neuropathic Symptoms and Signs

tribution of laparoscopic hernia repair group 86.8% (n=165) of them were male and 13.2% (n=25) of them were female. Regarding these demographic findings, no significance has been achieved in between the groups in gender difference.

Within the scope of the study, a total of 404 patients were included in the evaluation, including 214 Lichtenstein and 190 laparoscopic inguinal hernia repair. The distribution of the demographic and clinical findings of the patients according to the operation types was given in Table 1. When the table was examined, a statistically significant difference was observed between the groups in terms of LANSS scores (P<0.05). It was determined that the mean age of the patients with Lichtenstein repair was 51, and the mean LANSS Score was 3. It was determined that the mean age of the patients with laparoscopic inguinal hernia repair was 49 and the mean LANSS Score was 1.

The distribution of demographic and clinical findings of the patients who underwent laparoscopic inguinal hernia repair according to the operation methods was given in Table 2. When the table is examined, no statistically significant difference was observed between all demographic and clinical findings of the patients who underwent TAPP and TEP methods (P>0.05).

When the patients were analyzed in terms of laparoscopic inguinal hernia methods 23% (n=44) of them were operated via TAPP, and 77% (n=146) of them were operated via TEP. The mean LANSS score of the TAPP group was 1±3 and the TEP group was 2±3 with no significance (P=0.640) among laparoscopic inguinal hernia methods. However, individuals in the Lichtenstein repair (n=214) had LANSS score of a mean 3±4, while this score was 1±3 in the laparoscopic inguinal hernia (P<0.001).

DISCUSSION

Over the past 30 years, laparoscopic inguinal hernia repair has become widely available, thanks to advances in videoscopic equipment and patches. The most commonly used methods today; TEP and TAPP repairs. The most important advantages of laparoscopy are that the anatomy of the posterior wall of the inguinal region can be revealed relatively easily and that laparoscopic repairs have similar results with open repairs in terms of recurrence and complications [17]. As two minimally invasive surgical techniques, TAPP and TEP methods were first used in 1994 for inguinal hernia repair by Tetik *et al*. [15]. Although there are studies comparing laparoscopic inguinal hernia repair with open methods, there are few studies comparing TEP and TAPP methods [16, 18].

Laparoscopic repair is primarily recommended in cases of recurrence and bilateral inguinal hernia, since tissue changes secondary to surgery are not expected in the posterior area and both inguinal areas can be dominated from the same trocar entrances in cases of recurrence after the anterior approach [19]. Laparoscopic approach also has other advantages such as smaller incisions leading to earlier recovery time, less pain after surgery, early mobilization and early return to daily activities [20].

In our study, consistent with these literature data, the LANSS score was statistically significantly lower in the laparoscopic repair group than in the Lichtenstein

Table 2. Distribution of demographic and clinical findings according to laparoscopic inguinal hernia methods

Characteristics (n=190)	TAPP (n=44)	TEP (n=146)	P value
Age, years	52±14	48±14	0.091
LANSS score	1±3	2±3	0.640
Gender			0.718
Male	37 (84.1%)	128 (87.7%)	
Female	7 (15.9%)	18 (12.3%)	

Data are shown as mean±standard deviation or n (%). LANSS=Leeds Assessment of Neuropathic Symptoms and Signs, TAPP=Transabdominal preperitoneal, TEP=Totally extraperitoneal

repair group. In addition, there was no significant difference in LANSS scores between the TEP and TAPP groups in patients who underwent laparoscopic repair.

In previous literature, the rate of conversion to open technique during laparoscopic repair has been reported as 2-3%. Complications such as seroma, hematoma, testicular ischemia or pain, hydrocele, epididymitis, orchitis and chronic pain can be seen in the postoperative period [21]. Studies comparing the laparoscopic approach and the open anterior approach for hernia repair did not show any difference in recurrence rates. It is also emphasized that the learning time of laparoscopic methods is longer compared to open repairs and the recurrence rates are the same when performed in experienced hands [22].

The TAPP method is performed by entering the peritoneal cavity. There is no area restriction and it requires less experience than TEP. TEP, on the other hand, is thought to require more experience and is a somewhat more difficult technique than TAPP, since it is performed in a limited and less familiar area. However, since abdominal cavity is not entered in TEP risks such as organ injury, adhesion and infection are less than TAPP [22]. In a study by Matsumoto *et al*. (2018), when recurrences were observed more than expected in laparoscopic hernia repair, training courses were given to surgeons and it was observed that the recurrence rates decreased to the expected levels [23]. In the meta-analysis conducted by Wu *et al*. [24], the recurrence rate was reported as 3.8% in 1310 patients who underwent the TAPP technique, and it was observed that there was no significant difference in terms of recurrence when compared with the open technique.

In clinical practice and in the literature, there is much debate about the superiority of TAPP over TEP. Especially in bilateral inguinal hernia defects, the ability to repair intraperitoneally without entering the preperitoneal region seems to be a potential advantage of TAPP over TEP [16, 21]. During laparoscopic repair, hernia was detected in the contralateral region in 11-25% of patients with unilateral hernia in the preoperative physical examination [16]. The most important disadvantage during the application of TAPP is that the peritoneal dissection and flap must be closed again during the hernia repair. There is no need for such closure in patients who have undergone TEP. After laparoscopic repair, the peritoneum should be reapproximated to prevent the intra-abdominal organs from coming into contact with the mesh. Various methods of peritoneal flap closure have been described, such as tacker, fibrin adhesives and suture closure [16, 25].

While most of the studies on laparoscopic fixation devices today have focused on hernia recurrence that may occur in the postoperative period, there are few studies on the quality of life and pain of patients after laparoscopic hernia repair. In recent studies, postoperative chronic pain and decreased quality of life are the most common postoperative complications (1-54%) after laparoscopic inguinal hernia repair [26]. Although not certain, some authors consider young, female patients, recurrent and bilateral hernias, presence of preoperative pain, operation time and the number of tackers used during laparoscopic surgery as risk factors for postoperative pain and quality of life [27].

Laparoscopic hernia repair is recommended in case of recurrence after bilateral inguinal hernias and anterior approach repair [28]. TAPP and TEP methods are two important laparoscopic repair methods of inguinal hernia. The main difference between the TAPP and TEP method is the access route to the preperitoneal space. Although the results of TAPP and TEP procedures are similar in many respects, some results may differ. These differences may affect the technique preference in patient subgroups. TEP method is more suitable in patients with intra-abdominal adhesions because of not entering the abdomen. Because of the advantage of abdominal exploration, the TAPP method may be more suitable for laparoscopic repair of strangulated hernias [19, 29, 30].

It has been reported that the learning curve for the laparoscopic TEP method is completed after 60 cases [19]. The long learning curve is one of the main reasons why some surgeons avoid using the TEP method and prefer the TAPP method [22]. The results obtained by two physicians who achieved success in inguinal hernia repair is associated with the recurrence rates observed in the long-term, and these rates have been reported in the range of 1-2% for the TEP method and 0-3% for the TAPP method [25]. Recurrence was observed in 2 of our patients during follow-ups. Varcus *et al*. [16] reported length of hospital stay was 2 days for patients who underwent surgery using the TAPP & TEP method. In the study of Sağıroğlu *et al*. [26], all patients were discharged on the 1st postoperative day.

In our study all patients were discharged on the 1st post-operative day.

Several tackers are used in laparoscopic hernia surgeries. On the contrary, we have used only one tacker in each patient and this might be elaborated as the reason for the less chronic pain in our cases.

Strengths and Limitations

The strengths of our study include the use of a single surgical and anesthesia team and the large number of patients. However, the limitation of our study can be identified as the preoperative database not defining the subgroups of patients with inguinal hernia (direct, indirect, mixed, and femoral) and the size of the defect. If the study continues with more comprehensive data, it may allow for determining the most appropriate surgical procedure for patients during the preoperative assessment.

CONCLUSION

In conclusion, the use of TEP and TAPP methods in laparoscopic hernia repair was compared with Lichtenstein repair within the scope of this study. The TEP and TAPP methods have lower pain in the postoperative period.

Ethics Approval and Consent to Participate

This study was approved by the Bursa Yüksek Ihtisas Training and Reserach Hospital Clinical Research Ethics Committee (Decision No: 2011-KAEK-25 2021/12-22; date: 29.12.2021). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Written informed consent was obtained from all individual participants included in the study.

Data Availability

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

Authors' Contribution

Study Conception: MFE; Study Design: MD; Supervision: MFE; Funding: MD; Materials: MD; Data Collection and/or Processing: BD; Statistical Analysis and/or Data Interpretation: FG; Literature Review: SA; Manuscript Preparation: MFE; and Critical Review: BD.

Conflict of Interest

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

Financing

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

Acknowledgments

The authors have no acknowledgments to declare.

Generative Artificial Intelligence Statement

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

Editor's Note

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

REFERENCES

1. Tigora A, Radu PA, Garofil DN, et al. Modern Perspectives on Inguinal Hernia Repair: A Narrative Review on Surgical Techniques, Mesh Selection and Fixation Strategies. J Clin Med. 2025 9;14(14):4875. doi: 10.3390/jcm14144875.
2. Some Remarks on Hernia: III. Bassini's Operation for Radical Cure of an Inguinal Hernia. Hospital (Lond 1886). 1908 Oct 24;45(1157):91.
3. Messaris E, Nicastri G, Dudrick SJ. Total extraperitoneal laparoscopic inguinal hernia repair without mesh fixation: prospective study with 1-year follow-up results. Arch Surg. 2010;145(4):334-338. doi: 10.1001/archsurg.2010.20.
4. Messenger DE, Aroori S, Vipond MN. Five-year prospective follow-up of 430 laparoscopic totally extraperitoneal inguinal hernia repairs in 275 patients. Ann R Coll Surg Engl. 2010;92(3):201-205. doi: 10.1308/003588410X12628812458455.

5. Belyansky I, Tsirline VB, Klima DA, Walters AL, Lincourt AE, Heniford TB. Prospective, comparative study of postoperative quality of life in TEP, TAPP, and modified Lichtenstein repairs. Ann Surg. 2011;254(5):709-714; discussion 714-715. doi: 10.1097/SLA.0b013e3182359d07.
6. Ger R. The management of certain abdominal herniae by intra-abdominal closure of the neck of the sac. Preliminary communication. Ann R Coll Surg Engl. 1982;64(5):342-344.
7. Rosenberg J, Baig S, Chen DC, Derikx J. Groin hernia. Nat Rev Dis Primers. 2025;11(1):47. doi: 10.1038/s41572-025-00631-4.
8. Treede RD, Rief W, Barke A, et al. A classification of chronic pain for ICD-11. Pain. 2015;156(6):1003-1007. doi: 10.1097/j.pain.0000000000000160.
9. Song JW, Wolf JS Jr, McGillicuddy JE, Bhangoo S, Yang LJ. Laparoscopic triple neurectomy for intractable groin pain: technical report of 3 cases. Neurosurgery. 2011;68(2 Suppl Operative):339-46; discussion 346. doi: 10.1227/NEU.0b013e3182114480.
10. Lange JF, Kaufmann R, Wijsmuller AR, et al. An international consensus algorithm for management of chronic postoperative inguinal pain. Hernia. 2015;19(1):33-43. doi: 10.1007/s10029-014-1292-y.
11. Voorbrood CE, Burgmans JP, Van Dalen T, et al. An algorithm for assessment and treatment of postherniorrhaphy pain. Hernia. 2015;19(4):571-577. doi: 10.1007/s10029-015-1387-0.
12. Bjurstrom MF, Nicol AL, Amid PK, Chen DC. Pain control following inguinal herniorrhaphy: current perspectives. J Pain Res. 2014;7:277-290. doi: 10.2147/JPR.S47005.
13. Sacaklidir R, Sanal-Toprak C, Yucel FN, Gunduz OH, Sencan S. The Effect of Central Sensitization on Interlaminar Epidural Steroid Injection Treatment Outcomes in Patients with Cervical Disc Herniation: An Observational Study. Pain Physician. 2022;25(6):E823-E829.
14. Jeppesen M, Tolstrup MB, Gögenur I. Chronic Pain, Quality of Life, and Functional Impairment After Surgery Due to Small Bowel Obstruction. World J Surg. 2016;40(9):2091-2097. doi: 10.1007/s00268-016-3616-9.
15. Tetik C, Arregui ME, Dulucq JL, et al. Complications and recurrences associated with laparoscopic repair of groin hernias. A multi-institutional retrospective analysis. Surg Endosc. 1994;8(11):1316-1322; discussion 1322-1323. doi: 10.1007/BF00188291.
16. Varcuş F, Duta C, Dobrescu A, Lazar F, Papurica M, Tarta C. Laparoscopic Repair of Inguinal Hernia TEP versus TAPP. Chirurgia (Bucur). 2016;111(4):308-312.
17. Chung SD, Huang CY, Wang SM, et al. Laparoendoscopic single-site totally extraperitoneal adult inguinal hernia repair: initial 100 patients. Surg Endosc. 2011;25(11):3579-3583. doi: 10.1007/s00464-011-1761-7.
18. Choi YY, Han SW, Bae SH, Kim SY, Hur KY, Kang GH. Comparison of the outcomes between laparoscopic totally extraperitoneal repair and prolene hernia system for inguinal hernia; review of one surgeon's experience. J Korean Surg Soc. 2012;82(1):40-44. doi: 10.4174/jkss.2012.82.1.40.
19. Choi YY, Kim Z, Hur KY. Learning curve for laparoscopic totally extraperitoneal repair of inguinal hernia. Can J Surg. 2012;55(1):33-36. doi: 10.1503/cjs.019610.
20. Sharma D, Yadav K, Hazrah P, Borgharia S, Lal R, Thomas S. Prospective randomized trial comparing laparoscopic transabdominal preperitoneal (TAPP) and laparoscopic totally extra peritoneal (TEP) approach for bilateral inguinal hernias. Int J Surg. 2015;22:110-117. doi: 10.1016/j.ijsu.2015.07.713.
21. Bansal VK, Krishna A, Ghosh N, Bittner R, Misra MC. Comparison TAPP vs. TEP: Which Technique Is Better? In: Bittner R, Köckerling F, Fitzgibbons R Jr, LeBlanc K, Mittal S, Chowbey P (editors). Laparo-endoscopic Hernia Surgery. Springer, Berlin, Heidelberg. 2018: pp. 151-170.
22. Schouten N, Elshof JW, Simmermacher RK, et al. Selecting patients during the "learning curve" of endoscopic Totally Extraperitoneal (TEP) hernia repair. Hernia. 2013;17(6):737-743. doi: 10.1007/s10029-012-1006-2.
23. Matsumoto S, Hayakawa T, Kawarada Y, et al. Proper training in laparoscopic hernia repair is necessary to minimize the rising recurrence rate in Japan. Asian J Endosc Surg. 2018;11(2):151-154. doi: 10.1111/ases.12433.
24. Wu JJ, Way JA, Eslick GD, Cox MR. Transabdominal Pre-Peritoneal Versus Open Repair for Primary Unilateral Inguinal Hernia: A Meta-analysis. World J Surg. 2018;42(5):1304-1311. doi: 10.1007/s00268-017-4288-9.
25. Kumar M, Jha AK, Kumar D, Hasan I, Kumar D. An Observational Study on the Comparative Analysis of Trans Abdominal Preperitoneal and Totally Extraperitoneal: A Single-center Experience of 105 Patients. Ann Afr Med. 2025 Jul 23. doi: 10.4103/aam.aam_183_24.
26. Sağıroğlu J, Özdemir T, Atak T, et al. Laparoscopic Total Extraperitoneal Inguinal Hernia Repair Without Mesh Fixation: Report of Early Outcomes. South Clin Ist Euras 2016;27(3):215-219. doi: 10.14744/scie.2017.64872.
27. Thölix AM, Kössi J, Grönroos-Korhonen M, Harju J. Laparoscopic inguinal hernia repair with self-fixated meshes: a randomized controlled trial. Surg Endosc. 2025;39(4):2425-2435. doi: 10.1007/s00464-025-11616-5.
28. Buia A, Stockhausen F, Hanisch E. Laparoscopic surgery: A qualified systematic review. World J Methodol. 2015;5(4):238-254. doi: 10.5662/wjm.v5.i4.238.
29. Doudakmanis C, Kolla C, Bouliaris K, Efthimiou M, Koukoulis GD. Laparoscopic bilateral inguinal hernia repair: Should it be the preferred technique? World J Methodol. 2022;12(4):193-199. doi: 10.5662/wjm.v12.i4.193.
30. HerniaSurge Group. International guidelines for groin hernia management. Hernia. 2018;22(1):1-165. doi: 10.1007/s10029-017-1668-x.

The European Research Journal 2025;11(5):885-895 Original Article

DOI: https://doi.org/10.18621/eurj.1694766 *Obstetrics and Gynaecology*

Early-onset versus late-onset fetal cerebral ventriculomegaly: Sonographic characteristics and neonatal outcomes

Ruken Dayanan, Ahmet Arif Filiz, Merve Ayas Özkan, Dilara Duygulu Bulan, Hatice Ayhan, Ecem Bakan, Gülşan Karabay, Şevki Çelen

Department of Obstetrics and Gynecology, Division of Perinatology, Ankara Etlik City Hospital, Ankara, Türkiye

ABSTRACT

Objectives: This study aimed to compare the prenatal ultrasound characteristics, pregnancy outcomes, and neonatal prognosis of early- and late-onset fetal ventriculomegaly (VM).

Methods: A retrospective analysis was conducted on 102 pregnant women diagnosed with fetal VM, categorized into early-onset (≤24 weeks) and late-onset (>24 weeks) groups. Maternal characteristics, ventricular dimensions, associated anomalies, pregnancy outcomes, and neonatal parameters were compared between the groups.

Results: Early-onset VM was significantly associated with progressive ventricular enlargement, bilateral involvement, and a higher prevalence of additional anomalies detected via ultrasonography (70.4% vs. 29.2%, P<0.001) and Magnetic resonance imaging (MRI) (35.4% vs. 16.7%, P=0.030). Prenatal ultrasound findings differed significantly between the groups; early-onset VM cases more frequently exhibited bilateral (72.2% vs. 39.6%, P=0.002) and asymmetric (46.3% vs. 31.2%, P=0.037) ventricular enlargement, while late-onset VM was more commonly isolated (70.8% vs. 29.6%, P<0.001) and unilateral (60.4% vs. 27.8%). Live birth rates were lower (55.6% vs. 85.4%, P=0.001), pregnancy termination rates were higher (44.4% vs. 14.6%, P=0.001) and chromosomal abnormalities were higher (16.7% vs. 8.3%, P=0.246) in early-onset cases. Additionally, Apgar scores at 1 and 5 minutes were significantly lower in the early-onset group (P=0.028 and P=0.042, respectively).

Conclusions: Early-onset VM is more frequently associated with ventricular progression and structural anomalies, leading to poorer pregnancy and neonatal outcomes. These findings highlight the importance of close prenatal monitoring, including detailed ultrasound, fetal MRI, and genetic evaluation, to guide clinical management and parental counseling. Future studies with long-term neurodevelopmental follow-up are needed to further refine risk stratification and optimize patient care.

Keywords: Fetal cerebral ventriculomegaly, prenatal ultrasound, fetal MRI, pregnancy outcome, congenital anomalies, isolated ventriculomegaly

Fetal cerebral ventriculomegaly (VM) is one of the central nervous system anomalies characterized by enlargement of the lateral ventricles during prenatal ultrasonography and is among the most common neurological pathologies detected prenatally [1]. The incidence in the general population is

Received: May 8, 2025 **Accepted:** August 15, 2025 **Available Online:** August 21, 2025 **Published:** September 4, 2025

How to cite this article: Dayanan R, Filiz AA, Ayas Özkan M, et al. Early-onset versus late-onset fetal cerebral ventriculomegaly: Sonographic characteristics and neonatal outcomes. Eur Res J. 2025;11(5):885-895. doi: 10.18621/eurj.1694766

Corresponding author: Ruken Dayanan, MD., Phone: +90 312 797 00 00, E-mail: rukendayanan@gmail.com

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

approximately 1.5-2/1000 pregnancies and is classified as mild, moderate, and severe [2]. A lateral cerebral ventricular atrium diameter of 10 mm or more is accepted as the diagnostic criterion. VM may occur as an isolated finding or may be associated with genetic anomalies, intrauterine infections, congenital central nervous system malformations or vascular events [3]. The prognosis of VM identified in the prenatal period varies depending on the degree of ventricular enlargement, time of onset, and associated anomalies [2].

Fetal VM cases are divided into two categories as early-onset and late-onset in the prenatal period. Early-onset VM refers to cases detected before 24 weeks of gestation, while late-onset VM refers to cases detected after 24 weeks [2]. In the literature, it is reported that there are significant differences between early and late onset VM cases in terms of clinical course, prenatal follow-up, and postnatal outcomes. Early-onset cases have a higher incidence of genetic abnormalities and associated structural anomalies, and more unfavorable outcomes in terms of neurodevelopmental prognosis. In contrast, late-onset VM cases are reported to have a milder course and may spontaneously regress in some cases [4]. However, it is also known that severe neurodevelopmental effects may occur in the postnatal period in cases with progressive ventricular enlargement [1].

The long-term prognosis of fetal VM in the postnatal period depends on the degree of ventricular enlargement, associated neurologic abnormalities and whether the ventricular enlargement is progressive [5]. Most cases of mild VM have a favorable neurodevelopmental prognosis, whereas severe VM is associated with a higher risk of neurodevelopmental problems such as cerebral palsy, motor and cognitive retardation. Therefore, cases of fetal VM should be carefully followed up both prenatally and postnatally [6]. Although early- and late-onset fetal VM differ in clinical course and prognosis, comparative data remain limited. Existing studies often have small sample sizes, inconsistent definitions, and lack detailed evaluation of ventricular dynamics. This study addresses these gaps by systematically comparing early- and late-onset VM in a single well-defined cohort, integrating ultrasound, Magnetic resonance imaging (MRI), and neonatal outcomes to inform prenatal counseling and management. This study aims to compare the prenatal sonographic spectrum and neonatal outcomes of early and late onset fetal cerebral VM cases. Identifying differences between early and late onset cases will contribute to improving prenatal counseling processes and optimizing clinical management strategies.

METHODS

This retrospective cohort study included pregnant women admitted to the Perinatology Clinic of Ankara Etlik City Hospital between December 2022 and December 2024 and diagnosed with fetal VM. Diagnosis and follow-up information were obtained retrospectively from the hospital information management system and prenatal and neonatal outcomes were evaluated. The diagnosis of fetal VM was based on prenatal ultrasonography with a lateral ventricular atrial diameter ≥10 mm. All ultrasound examinations had been performed as part of routine prenatal care using a Voluson E10 system (GE Healthcare, Chicago, IL, USA) equipped with a 2-5 MHz convex transducer. Measurements were obtained transabdominally at the level of the lateral ventricular atrium in a standard coronal plane. The inclusion criteria for the study were singleton pregnancies diagnosed with VM based on an atrial width of 10 mm or greater, and with at least one follow-up ultrasound examination available in the records. Pregnancies were excluded if they were multiple gestations, lacked follow-up data, or were complicated by major maternal systemic diseases such as pregestational diabetes or chronic hypertension. The diagnosed cases were classified as early or late onset VM according to gestational week. Prenatal findings, neonatal outcomes and perinatal morbidity and mortality rates were compared between the groups.

A total of 102 pregnant women were included in the study. Patients were divided into two groups according to the gestational week at which fetal VM was diagnosed. The early onset group (n=54) consisted of patients diagnosed before 24 weeks of gestation, and the late onset group (n=48) consisted of patients diagnosed after 24 weeks of gestation.

Within the scope of detailed ultrasound evaluation, whether VM was symmetric or asymmetric, whether it showed unilateral or bilateral enlargement, and dynamic changes in ventricular width were recorded. Asymmetry was defined as a ≥2 mm difference in width between the lateral ventricles. Unilateral VM was defined as ≥10 mm diameter of only one lateral

ventricle, whereas bilateral VM was defined as ≥10 mm diameter of both lateral ventricles.

It was also monitored whether the VM was transient or permanent (progressive). Transient VM was defined as a decrease in ventricular width to <10 mm during follow-up. In contrast, persistence of ventricular width ≥10 mm on multiple ultrasound assessments during follow-up was classified as persistent VM. A change of ≤2 mm in ventricular diameter compared with the initial measurement indicates stability, an increase of >2 mm indicates progression, and a decrease in ventricular diameter to <10 mm is considered spontaneous regression. Isolated VM (IVM) was defined as a special category and diagnosed in the absence of associated structural anomalies, chromosomal abnormalities or intrauterine infections. Cases with associated anomalies are classified as non-isolated VM (non-IVM).

In the prenatal period, the isolated VM, presence of additional structural anomalies, bilateral or unilateral enlargement, ventricular enlargement dynamics (progression, regression, or stability) and TORCH test results were analyzed. In the neonatal period, pregnancy outcomes, birth week, birth weight, Apgar scores, need for neonatal intensive care, perinatal morbidity and mortality rates were evaluated. Composite adverse perinatal outcomes (CAPOs) include the presence of at least one of the following adverse outcomes: transient tachypnea of the newborn, respiratory distress syndrome, need for continuous positive airway pressure, need for mechanical ventilation, need for phototherapy, neonatal hypoglycemia, intraventricular hemorrhage, necrotizing enterocolitis, neonatal sepsis, 5th minute APGAR score <7, neonatal intensive care unit (NICU) admission, placental abruption, and preterm birth.

This study was conducted in compliance with the Declaration of Helsinki and was approved by the Ankara Etlik City Hospital Scientific Research Evaluation and Ethics Committee (decision number: AEŞH-BADEK-2025-0259). In this study, due to its retrospective nature, informed consent was waived with the approval of the Ethics Committee. All data were anonymized, and participant confidentiality was strictly maintained.

Statistical Analysis

Statistical analysis was performed using IBM Corporation SPSS version 22.0 (IBM Corporation, Armonk, NY, USA). The Kolmogorov-Smirnov test was used to analyze conformity to normal distribution. Descriptive statistics of continuous variables are shown as "mean ± standard deviation" for those with a normal distribution. and as “median (interquartile range)” for those that do not. Continuous variables that were and were not normally distributed were compared using the Mann-Whitney U test. Normally distributed data were compared using the independent sample t-test. Statistical significance for the tests was defined as a P-value of less than 0.05.

RESULTS

A total of 102 pregnant women were included in the study, and the cases were divided into two groups as early onset (≤24 weeks) and late onset (>24 weeks) according to the gestational week at which fetal VM was detected. There was no significant difference between the groups in terms of maternal age and body mass index (BMI), but the number of gravida was higher in the early-onset group (P=0.036). TORCH test results were similar, but the rate of additional anomaly detection by MRI was significantly higher in the early-onset group (35.4% vs. 16.7 %, P=0.030). The diagnosis was made at a median of 21 weeks in the early-onset group and 28 weeks in the late-onset group, and this difference was statistically significant (P<0.001). There was no significant difference between the groups in terms of ventricular width (P=0.443). Cytogenetic testing was performed in 53.7% of the early-onset group and 29.2% of the late-onset group, with a statistically significant difference between the groups (P=0.021). Regarding chromosomal abnormalities, 16.7% of cases in the early-onset group tested positive, compared to 8.3% in the late-onset group. However, this difference was not statistically significant (P=0.246) (Table 1).

The perinatal ultrasound characteristics of VM differed significantly between the early-onset and late-onset groups. The incidence of isolated VM was significantly higher in the late-onset group compared to the early-onset group (70.8% vs. 29.6%, P<0.001). Conversely, non-isolated VM was more prevalent in the early-onset group (70.4% vs. 29.2%). Laterality also demonstrated a significant difference between the

Table 1. Characteristics of the study population

	Early-onset group (n=54)	Late-onset group (n=48)	P value
Maternal age (year)			0.178[a]
<35	6 (11.1%)	10 (20.8%)	
>35	48 (88.9%)	38 (79.2%)	
BMI (kg/m$^{2)}$	31.4±17.2	29.8±6.3	0.528[b]
Gravida	2 (3)	2 (2)	**0.036[c]**
Parity	1 (2)	1 (1)	0.058[c]
Nulliparous	17 (31.5%)	24 (50%)	0.057[a]
In vitro fertilization	0 (0%)	0 (0%)	NA
Gestational age at initial diagnosis (weeks)	21 (3)	28 (5)	**<0.001[c]**
Ventricular size (mm)	13.6±6.1	12.9±4.1	0.443[b]
Additional abnormalities on MRI	17 (35.4%)	9 (16.7%)	**0.030[a]**
Cytogenetic testing			**0.021[a]**
Absent	25 (46.3%)	34 (70.8%)	
Present	29 (53.7%)	14 (29.2%)	
Chromosomal abnormalities			0.246[d]
Negative	45 (83.3%)	44 (91.7)	
Positive	9 (16.7%)	4 (8.3)	
TORCH			0.552[a]
Negative	52 (96.3%)	45 (93.8%)	
Positive	2 (3.7%)	3 (6.3%)	

Data are shown as mean±standard deviation or median (interquartile range) or n (%) where appropriate. BMI=Body mass index, TORCH=Toxoplasma gondii, other, rubella virus, cytomegalovirus and herpes simplex virus, MRI=Magnetic resonance imaging, NA=Not applicable.

P<0.05 indicates a significant difference and statistically significant P-values are in bold

[a]Pearson chi-square, [b]Student t-test, [c]Mann Whitney-U test, [d]Fisher's exact test

groups, with unilateral VM being more frequent in the late-onset group (60.4% vs. 27.8%), while bilateral VM was more common in the early-onset group (72.2% vs. 39.6%, P=0.002). Similarly, asymmetric VM was more prevalent in the late-onset group (68.8% vs. 46.3%), whereas the early-onset group had a higher proportion of symmetric VM (53.7% vs. 31.2%, P=0.037). Regarding intrauterine changes, intrauterine regression was more commonly observed in the late-onset group (29.2% vs. 16.7%), while intrauterine progression was higher in the early-onset group (46.3% vs. 60.4%). Additionally, intrauterine stabilization was significantly more frequent in the early-onset group (37% vs. 10.4%, P=0.007). No significant differences were found between the groups in terms of transience (P=0.092), severity of VM (P=0.547), or amniotic fluid status (P=1.000). These findings suggest that early-onset VM is more frequently associated with non-isolated, bilateral, and symmetric forms, whereas late-onset VM is more likely to be isolated, unilateral, and asymmetric (Table 2).

Within the scope of the study, a total of 31 pregnant women accepted pregnancy termination. In terms of pregnancy outcomes, the termination rate was significantly higher in the early-onset group (44.4% vs. 14.6%, P=0.001), while the live birth rate was significantly higher in the late-onset group (85.4% vs. 55.6%). Although there was no significant difference

Table 2. Perinatal ultrasound characteristics of VM in the early-onset and late-onset groups

	Early-onset group (n=54)	Late-onset group (n=48)	P value
Isolated or not			**<0.001[a]**
Isolated	16 (29.6%)	34 (70.8%)	
Non-isolated	38 (70.4%)	14 (29.2%)	
Laterality			**0.002[a]**
Unilateral	15 (27.8%)	29 (60.4%)	
Bilateral	39 (72.2%)	19 (39.6%)	
Symmetry			**0.037[a]**
Symmetric	29 (53.7%)	15 (31.2%)	
Asymmetric	25 (46.3%)	33 (68.8%)	
Transience			0.092[a]
Transient	13 (24.1%)	19 (39.6%)	
Non-transient	41 (75.9%)	29 (60.4%)	
Intrauterine changes			**0.007[a]**
Intrauterine regression	9 (16.7%)	14 (29.2%)	
Intrauterine progression	25 (46.3%)	29 (60.4%)	
Intrauterine stabilization	20 (37%)	5 (10.4%)	
Severity of VM			0.547[a]
Mild	28 (51.9%)	30 (62.5%)	
Moderate	8 (14.8%)	6 (12.5%)	
Severe	18 (33.3%)	12 (25%)	
Amnion fluid			1[a]
Normal	48 (88.9%)	44 (91.7%)	
Oligohydramnios	1 (1.9%)	0 (0%)	
Polyhydramnios	5 (9.3%)	4 (8.3%)	

Data are shown as n (%). VM=Ventriculomegaly

P<0.05 indicates a significant difference and statistically significant P-values are in bold.

[a]Pearson chi-square

between the groups in terms of gestational week, prematurity rate, birth weight and cesarean section rates (P>0.05). Apgar scores at the 1st and 5th minute were significantly higher in the late-onset group (P=0.028 and P=0.042) (Table 3).

When additional anomalies detected by prenatal ultrasonography were compared, the proportion of cases without additional anomalies was significantly higher in the late-onset group (70.8% vs. 29.6%, P<0.001). Central nervous system (CNS) anomalies were more common in the early-onset group (44.4% vs. 18.8%), but there was no significant difference between the groups in terms of non-CNS anomalies (7.4% vs. 6.3%, P>0.05). However, multisystem anomalies were more common in the early-onset group (18.5% vs. 4.2%), suggesting that early-onset cases may be associated with more complex fetal anomalies (Table 4).

DISCUSSION

In this study, the effects of fetal VM on prenatal ultrasonographic findings, neonatal outcomes and perinatal

Table 3. Pregnancy outcomes

	Early-onset group (n=54)	Late-onset group (n=48)	P value
Pregnancy outcomes			**0.001[a]**
Live birth	30 (55.6%)	41 (85.4%)	
Termination of pregnancy	24 (44.4%)	7 (14.6%)	
Live birth analysis	**n= 30**	**n=41**	
Gestational age at delivery (week)	38 (2)	38 (2)	0.421[b]
Prematurity (<37 weeks)	5 (16.7%)	5 (12.2%)	0.512[a]
Birth weight (gram)	2720±966	3014±555	0.111[c]
Cesarean section	20 (66.7%)	27 (65.9%)	0.943[a]
Apgar score at 1st minute	8 (3)	9 (1)	**0.028[b]**
Apgar score at 5th minute	9 (2)	10 (1)	**0.042[b]**
CAPO	9 (30%)	10 (24.4%)	0.538[a]
NICU admission	8 (26.7%)	9 (22%)	0.531[a]
Transient tachypnea of the newborn	4 (13.3%)	1 (2.4%)	0.151[d]
Neonatal sepsis	0 (0%)	0 (0%)	NA
Respiratory distress syndrome	5 (16.7%)	4 (9.8%)	0.471[d]
Continuous positive airway pressure	4 (13.3%)	3 (7.3%)	0.435[d]
Mechanical ventilation	2 (6.7%)	2 (4.9%)	1[d]
Phototherapy for neonates	4 (13.3%)	1 (2.4%)	0.151[d]
Necrotizing enterocolitis	0 (0%)	1 (2.4%)	1[d]
Intraventricular hemorrhage	0 (0%)	0 (0%)	NA
Seizure history	1 (3.3%)	0 (0%)	0.412[d]

Data are shown as mean±standard deviation or median (interquartile range) or n (%) where appropriate. CAPO=Composite adverse perinatal outcome, NICU=Neonatal intensive care unit, NA=Not applicable.

P<0.05 indicates a significant difference and statistically significant P-values are in bold

[a]Pearson chi-square, [b]Student t-test, [c]Mann Whitney-U test, [d]Fisher's exact test

prognosis were compared between early and late onset groups. In this study, we found that early-onset fetal VM was more frequently associated with additional anomalies, bilateral, symmetric involvement, and progressive ventricular enlargement, whereas late-onset VM was more commonly isolated, unilateral, asymmetric and had a higher likelihood of spontaneous regression. The higher rate of chromosomal abnormalities and MRI-detected additional anomalies in the early-onset group supports the hypothesis that early-onset VM is often part of a broader pathological process rather than an isolated finding. Furthermore, pregnancy termination was significantly more common in the early-onset group, while live birth rates were higher in late-onset cases, suggesting a better perinatal prognosis for late-onset VM. Additionally, Apgar scores at both the 1st and 5th minutes were significantly higher in the late-onset group, reinforcing the association between early-onset VM and poorer neonatal outcomes. These findings highlight the importance of distinguishing between early- and late-onset VM in clinical practice to better predict prenatal progression and neonatal prognosis.

Previous studies suggest that early- and late-onset VM have distinct etiological and clinical courses. Early-onset fetal VM has been associated with a higher prevalence of structural and chromosomal anomalies, contributing to a less favorable perinatal

Table 4. Associated abnormalities found on prenatal ultrasound

	Early-onset group (n=54)	Late-onset group (n=48)	P value
Absent	16 (29.6%)	34 (70.8%)	**<0.001**[a]
CNS	24 (44.4%) Agenesis of the corpus callosum (n=5) Dysgenesis of the corpus callosum (n=2) Open spina bfida (n=11) Z-shaped brain stem (n=1) Dandy Walker malformation (n=1) Vein of Galen aneurysmal (n=1) Dural sinus thrombosis (n=1) Interhemispheric cyst (n=1) Hypoplastic cerebellum (n=1)	9 (18.8%) Agenesis of the corpus callosum (n=4) Open spina bfida (n=4) Intraventricular hemorrhage (n=1)	
Non-CNS	4 (7.4%) Right aortic arch (n=1) Single umbilical artery (n=1) Diaphragmatic hernia (n=1) ARSA (n=1)	3 (6.3%) HL-FL shortening (n=1) Diaphragmatic hernia (n=1) Bilateral pyelectasis (n=1)	
Multiple systems	10 (18.5%) Bilateral renal agenesis and diastematomyelia (n=1) Midline syndactyly and cranial irregularities (n=1) Megacystis, rhombencephalosynapsis and single umbilical artery (n=1) Bilateral pelviectasis+cleft lip and palate (n=1) Unilateral pyelectasis+ inlet VSD+ right aortic arch (n=1) Cleft lip, absent stomach, hypoplastic cerebellum, pericardial effusion, low-set ears (n=1) Lymphatic drainage disorder, Rocker bottom, syndactyly (n=1) Corpus callosum dysgenesis, VSD, Mitral atresia, persistent left superior vena cava, aortic hypoplasia (n=1) Bilateral pes equinovarus, small bladder, pericardial effusion (n=1) Absent CSP, cystic hygroma, skin edema, microphthalmia, pleural effusion, two arteries observed, absent stomach, clenched hands, bilateral pes equinovarus (n=1)	2 (4.2%) Diaphragmatic hernia, Tetralogy of Fallot, right atrial aneurysm, cleft lip/palate (n=1) Dysmorphic corpus callosum, cleft lip, hypertelorism, congenital hepatic fibrosis, generalized ascites, autosomal recessive polycystic kidney disease, severe micromelia, fetal dextrocardia, inlet VSD, hypoplastic aorta (n=1)	

Data are expressed as n (%). ARSA=Aberrant right subclavian artery, CNS=Central nervous system, CSP=Cavum septi pellucidi, FL=femur length, HL=humerus length, VSD=ventricular septal defect

prognosis [4]. Consistent with this, our study found that additional anomalies were significantly more prevalent in early-onset cases compared to late-onset cases (70.4% vs. 29.2%, P<0.001). In a large-scale study, Bhatia *et al*. reported that aneuploidy, agenesis of the corpus callosum, spina bifida, and Dandy-Walker malformation were among the most common anomalies in early-onset VM, whereas late-onset cases were more often isolated and associated with aqueductal stenosis, cerebral hemorrhage, and porencephaly [2]. In another study examining the etiology of fetal VM in detail, Pappas *et al*. [7] reported that early-onset cases were generally associated with neurodevelopmental malformations and chromosomal anomalies, whereas late-onset VM was mostly isolated and had a better prognosis. Our findings align with these results, reinforcing that early-onset VM is frequently part of a broader pathological process, whereas late-onset VM is more likely to be an isolated, potentially regressive condition.

Wang *et al*. [4] conducted a study comparing the ultrasonographic characteristics of early- and late-onset VM cases and reported significant differences between the two groups. According to their findings, isolated, unilateral, asymmetric involvement is more common in late-onset VM, whereas bilateral, symmetric involvement and additional structural anomalies are more common in early-onset cases. Our study is largely consistent with these results. The rate of bilateral and symmetric involvement was significantly higher in early-onset cases (P=0.037), whereas late-onset VM cases were more often unilateral and asymmetric. Furthermore, Wang *et al*. [4] showed that the rate of intrauterine progression was significantly higher (P=0.03) in early-onset mild VM cases. It has also been reported in the literature that early-onset VM is more likely to progress and may have a more unfavorable neurodevelopmental prognosis [8]. Similarly, in our study, it was found that ventricular enlargement tended to progress in early-onset VM, and this increased poor pregnancy outcomes.

In our study, the rates of additional anomalies in the early-onset VM group were significantly higher by both ultrasonography (70.4% vs. 29.2%, P<0.001) and MRI (35.4% vs. 16.7%, P=0.030). As reported in the literature, MRI is a superior method compared to ultrasonography in detecting corpus callosum anomalies, posterior fossa defects, and migration disorders [9, 10]. The most common CNS anomalies in early-onset cases were agenesis of the corpus callosum, open spina bifida, and brainstem anomalies, suggesting that early-onset VM may be part of a broader spectrum of neurodevelopmental disorders rather than an isolated finding. Similarly, multisystem anomalies affecting the renal, cardiac, and skeletal systems were more frequent in early-onset cases, with severe structural abnormalities such as bilateral renal agenesis, diastematomyelia, and megacystis, further supporting the poor fetal prognosis in this group. Consistent with our findings, Wang *et al*. [4] also reported that CNS and multisystem anomalies were significantly more prevalent in the early-onset VM group compared to late-onset cases. These similarities between studies further emphasize the importance of comprehensive prenatal evaluation, including MRI and genetic testing, in early-onset VM cases to better predict prognosis and optimize perinatal management.

In addition, in a large-scale study by Carta *et al*. [11], it was emphasized that neurodevelopmental prognosis was generally favorable in isolated cases of VM, but cases with large ventricular diameter (>15 mm) or progression carried a higher risk. Another important study by Ali *et al*. [12] examined the long-term neurodevelopmental outcomes of fetal VM cases. In this study, it was reported that the majority of isolated and mild VM cases had normal neurodevelopmental outcomes, but progressive VM was associated with severe motor and cognitive sequelae. Similarly, our study demonstrated that early-onset VM cases exhibited a greater tendency for ventricular enlargement and progression over time. This progressive nature of early-onset VM was associated with poorer pregnancy outcomes, including significantly lower live birth rates (55.6% vs. 85.4%, P=0.001) and a higher frequency of pregnancy termination compared to late-onset cases. Furthermore, neonatal outcomes were also worse in the early-onset group, as reflected by significantly lower Apgar scores at both the 1st and 5th minutes. These findings highlight the importance of monitoring ventricular dynamics in early-onset VM, as progressive ventricular dilation may influence clinical decision-making and pregnancy management. These data suggest that, especially late-onset and isolated cases, offer a better prognosis, and follow-up protocols should be individualized accordingly.

VM is associated with most common chromoso-

mal abnormalities, particularly trisomy 21, trisomy 18, and sex chromosome anomalies [13]. The frequency of chromosomal anomaly in cases of fetal VM varies depending on factors including the degree of ventricular enlargement, the presence of additional structural anomalies, and the onset time of VM. In the literature, while the rate of chromosomal anomaly varies between 3-5% in isolated cases of VM, this rate increases up to 17-20% in cases accompanied by additional anomalies [14]. Bhatia *et al*. [2] reported that chromosomal abnormalities were more common in the early-onset VM group and the aneuploidy rate in this group was 13.4%. Wang *et al*. [4] also supported these findings and reported that early-onset VM was generally associated with genetic syndromes, whereas late-onset VM cases had mostly normal chromosomal structure. In our study, the rate of chromosomal anomaly was found to be higher in the early-onset VM group (16.7% vs. 8.3%), but the difference was not statistically significant (P=0.246). In terms of prenatal genetic evaluation, cytogenetic testing and other genetic tests are recommended, especially in cases of VM accompanied by additional anomalies or showing progression. In our study, the rate of cytogenetic testing was found to be significantly higher in the early-onset VM group (53.7% vs. 29.2%, P=0.021), suggesting that this group should be evaluated more in terms of genetic analysis.

The decision to terminate pregnancies with fetal VM varies depending on the degree of ventricular enlargement, progression, presence of additional anomalies and chromosomal abnormalities [15]. In the literature, it has been reported that pregnancy termination rates are higher in cases of severe VM (>15 mm) and progressive VM, whereas isolated and mild cases are more frequently continued [16]. In a study by Chervenak *et al*. [17], 26% of pregnancies diagnosed with VM resulted in elective abortion. In another study by Vintzileos *et al*. [18], 20% induced abortion was reported in similar cases. It has been emphasized that termination rates are higher, especially in cases of VM associated with early onset and multisystem anomalies. In our study, the termination rate was significantly higher in the early-onset VM group (44.4% vs. 14.6%, P=0.001). This finding suggests that early-onset VM is usually associated with more severe fetal pathologies and families may be more prone to termination decision in this situation. Similarly, Wang *et al*. [4] reported that pregnancy termination rates were higher in the early-onset VM group, whereas late-onset and isolated cases were more frequently continued. These findings emphasize that the decision for termination should be based not only on ventricular size but also on the presence of additional anomalies, progression in ventricular diameter, and genetic outcomes. It is especially important to evaluate VM cases with early onset and associated with multisystem anomalies more carefully and to inform families in detail during prenatal counseling.

The less favorable prognosis observed in early-onset VM may be explained by its occurrence during critical stages of fetal neurodevelopment, when neuronal proliferation, migration, and cortical organization are highly active. Disruption of these processes in the first half of gestation is often linked to underlying genetic or complex structural anomalies, which may affect multiple organ systems [11]. In contrast, late-onset VM generally develops after key neurodevelopmental milestones, frequently due to mechanical obstruction of cerebrospinal fluid pathways or mild intraventricular hemorrhage, and is therefore more likely to be isolated, non-progressive, and associated with a favorable prognosis [2]. These mechanistic differences have direct implications for prenatal counseling and management. In early-onset VM, the higher likelihood of associated anomalies and ventricular progression justifies a more intensive monitoring strategy, including comprehensive neurosonography, fetal MRI, and genetic evaluation. Conversely, late-onset isolated cases, particularly those showing stability or regression, can be managed with less intensive surveillance, and counseling can emphasize their generally favorable outlook. By tailoring follow-up and counseling strategies to the timing of onset, ventricular dynamics, and associated findings, clinicians can provide more accurate prognostic information and optimize perinatal outcomes.

Strengths and Limitations

Building on these observations, our study provides a comprehensive evaluation of early- and late-onset fetal VM, comparing their prenatal ultrasound characteristics, pregnancy outcomes, and neonatal prognosis. One of the key strengths of our study is its detailed assessment of ventricular dynamics, emphasizing the progressive nature of early-onset VM, which has been less frequently highlighted in previous research. The

use of both prenatal ultrasonography and fetal MRI allowed for a more precise evaluation of associated anomalies, enhancing the reliability of our findings. Furthermore, by comparing pregnancy outcomes between the two groups, our study contributes valuable clinical insights into the management of VM, particularly in cases with progressive ventricular enlargement.

However, some limitations should be acknowledged. First, this study was conducted in a single center, which may limit the generalizability of the findings to broader populations. Additionally, while we analyzed the association between VM and chromosomal abnormalities, not all cases underwent invasive genetic testing, which may have affected the identification of underlying genetic conditions. Another limitation is that our study primarily focuses on prenatal findings and short-term neonatal outcomes; therefore, long-term neurodevelopmental follow-up is needed to further assess the prognostic implications of early- versus late-onset VM. Despite these limitations, our study provides clinically significant findings that reinforce the importance of differentiating early- and late-onset VM in prenatal counseling and pregnancy management. Future studies with larger, multi-center cohorts and long-term follow-up data would further clarify the impact of ventricular progression on neurodevelopmental outcomes and refine clinical management strategies.

CONCLUSION

This study highlights the clinical differences between early- and late-onset fetal VM. Early-onset VM was more frequently associated with progressive ventricular enlargement, bilateral involvement, and additional anomalies, whereas late-onset VM was more commonly isolated with a higher likelihood of spontaneous resolution. Pregnancy outcomes were also poorer in early-onset cases, emphasizing the need for closer prenatal monitoring, including detailed ultrasound, fetal MRI, and genetic evaluation. Future studies with long-term follow-up are needed to further refine clinical management strategies.

Ethics Approval and Consent to Participate

This study was approved by the Ankara Etlik City Hospital Scientific Research Evaluation and Ethics Committee (Decision No: AEŞH-BADEK-2025-0259; date: 26.03.2025). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Informed consent was waived because of the retrospective nature of the study. All data were anonymized, and participant confidentiality was strictly maintained.

Data Availability

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

Authors' Contribution

Study Conception: RD; Study Design: RD; Supervision: SC; Funding: N/A; Materials: GK; Data Collection and/or Processing: MAO, DDB, HA, EB; Statistical Analysis and/or Data Interpretation: AAF; Literature Review: RD; Manuscript Preparation: RD; and Critical Review: RD, SC.

Conflict of Interest

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

Financing

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

Acknowledgments

The author(s) have no acknowledgments to declare.

Generative Artificial Intelligence Statement

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

Editor's Note

All statements made in this article are solely those of the authors and do not represent the views of their

affiliates or the publisher, editors, or reviewers. Any claims made by any product or manufacturer that may be evaluated in this article are not guaranteed or endorsed by the publisher.

REFERENCES

1. Society for Maternal-Fetal Medicine (SMFM); Fox NS, Monteagudo A, Kuller JA, Craigo S, Norton ME. Mild fetal ventriculomegaly: diagnosis, evaluation, and management. Am J Obstet Gynecol. 2018;219(1):B2-B9. doi: 10.1016/j.ajog.2018.04.039.
2. Bhatia A, Thia EWH, Bhatia A, Ruochen D, Yeo GSH. Sonographic spectrum and postnatal outcomes of early-onset versus late-onset fetal cerebral ventriculomegaly. J Matern Fetal Neonatal Med. 2022;35(23):4612-4619. doi: 10.1080/14767058.2020.1857358.
3. Giorgione V, Haratz KK, Constantini S, Birnbaum R, Malinger G. Fetal cerebral ventriculomegaly: What do we tell the prospective parents? Prenat Diagn. 2022;42(13):1674-1681. doi: 10.1002/pd.6266.
4. Wang X, Zhang S, Wang J, Zhang S, Feng L, Wu Q. Follow-up outcome analysis of 324 cases of early-onset and late-onset mild fetal ventriculomegaly: a retrospective cohort study. Eur J Med Res. 2024;29(1):128. doi: 10.1186/s40001-024-01709-7.
5. Rosseau GL, McCullough DC, Joseph AL. Current prognosis in fetal ventriculomegaly. J Neurosurg. 1992;77(4):551-555. doi: 10.3171/jns.1992.77.4.0551.
6. Winkler A, Tölle S, Natalucci G, Plecko B, Wisser J. Prognostic Features and Long-Term Outcome in Patients with Isolated Fetal Ventriculomegaly. Fetal Diagn Ther. 2018;44(3):210-220. doi: 10.1159/000480500.
7. Pappas A, Adams-Chapman I, Shankaran S, et al; Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Neurodevelopmental and Behavioral Outcomes in Extremely Premature Neonates With Ventriculomegaly in the Absence of Periventricular-Intraventricular Hemorrhage. JAMA Pediatr. 2018;172(1):32-42. doi: 10.1001/jamapediatrics.2017.3545.
8. Horgos B, Mecea M, Boer A, et al. White matter changes in fetal brains with ventriculomegaly. Front Neuroanat. 2023;17:1160742. doi: 10.3389/fnana.2023.1160742.
9. Kyriakopoulou V, Vatansever D, Elkommos S, et al. Cortical overgrowth in fetuses with isolated ventriculomegaly. Cereb Cortex. 2014;24(8):2141-2150. doi: 10.1093/cercor/bht062.
10. Di Mascio D, Sileo FG, Khalil A, et al. Role of magnetic resonance imaging in fetuses with mild or moderate ventriculomegaly in the era of fetal neurosonography: systematic review and meta-analysis. Ultrasound Obstet Gynecol. 2019;54(2):164-171. doi: 10.1002/uog.20197.
11. Carta S, Kaelin Agten A, Belcaro C, Bhide A. Outcome of fetuses with prenatal diagnosis of isolated severe bilateral ventriculomegaly: systematic review and meta-analysis. Ultrasound Obstet Gynecol. 2018;52(2):165-173. doi: 10.1002/uog.19038.
12. Ali F, Gurung F, Nanda S, et al. Perinatal and neurodevelopmental outcomes of fetal isolated ventriculomegaly: a systematic review and meta-analysis. Transl Pediatr. 2024;13(4):555-574. doi: 10.21037/tp-23-548.
13. Zhao D, Cai A, Wang B, Lu X, Meng L. Presence of chromosomal abnormalities in fetuses with isolated ventriculomegaly on prenatal ultrasound in China. Mol Genet Genomic Med. 2018;6(6):1015-1020. doi: 10.1002/mgg3.477.
14. Huang RN, Chen JY, Pan H, Liu QQ. Correlation between mild fetal ventriculomegaly, chromosomal abnormalities, and copy number variations. J Matern Fetal Neonatal Med. 2022;35(24):4788-4796. doi: 10.1080/14767058.2020.1863941.
15. Chu N, Zhang Y, Yan Y, Ren Y, Wang L, Zhang B. Fetal ventriculomegaly: Pregnancy outcomes and follow-ups in ten years. Biosci Trends. 2016;10(2):125-132. doi: 10.5582/bst.2016.01046.
16. Hudgins RJ, Edwards MS, Goldstein R, et al. Natural history of fetal ventriculomegaly. Pediatrics. 1988;82(5):692-697.
17. Chervenak FA, Duncan C, Ment LR, et al. Outcome of fetal ventriculomegaly. Lancet. 1984;2(8396):179-181. doi: 10.1016/s0140-6736(84)90477-x.
18. Vintzileos AM, Campbell WA, Weinbaum PJ, Nochimson DJ. Perinatal management and outcome of fetal ventriculomegaly. Obstet Gynecol. 1987;69(1):5-11.

The European Research Journal 2025;11(5):896-906 Original Article

DOI: https://doi.org/10.18621/eurj.1653062 *Public Health*

Associations between anxiety, depression, and sleep problems in women during the COVID-19 pandemic: a cross sectional study

Şeyma Çatalgöl[1], Rabia Akgün[2]

[1]Department of Nursing, Uşak University, Faculty of Health Sciences, Uşak, Türkiye; [2]Dokuz Eylul University, Research and Application Hospital, İzmir, Türkiye

ABSTRACT

Objectives: This study was conducted to determine the association between anxiety, depression, and sleep problems in women during the COVID-19 pandemic.

Methods: This cross-sectional study included 686 women via a web survey between February 20, 2021, and April 9, 2021. In this study, participants' anxiety scores were evaluated with the Beck Anxiety Inventory (BAI), depression findings with the Beck Depression Inventory (BDI), and sleep quality scores with the Pittsburgh Sleep Quality Index (PSQI). Cut-off points: PSQI >5 indicates poor sleep quality, BAI 16-25 indicates moderate, and ≥26 indicates severe anxiety; BDI 17-29 indicates moderate, and ≥30 indicates severe depression. Statistical analyses included t-test, ANOVA, Pearson correlation, and multiple linear regression.

Results: The younger age group, students, singles, smokers, and alcohol users were at higher risk for increased anxiety, depression, and poor sleep quality during the COVID-19 pandemic. Participants exhibited high levels of anxiety (moderate anxiety in 20.12%, severe anxiety in 20.70%), depression (moderate depression in 25.22%, severe depression in 12.53%), and poor sleep quality (58.6%). A moderate positive correlation was found between sleep quality and anxiety ($r=0.517$, $P<0.01$) and depression ($r=0.513$, $P< 0.01$). A strong positive correlation was observed between anxiety and depression ($r=0.647$, $P<0.01$). Multiple linear regression analysis revealed that anxiety ($\beta=0.071$, $P<0.001$) and depression ($\beta=0.075$, $P<0.001$) were significant predictors of sleep quality.

Conclusions: The findings show that the COVID-19 pandemic negatively affected women's mental health and sleep quality. Anxiety and depression significantly impact sleep quality, and early intervention, long-term follow-up, and national-level measures are necessary to prevent the complications of psychological issues and sleep disorders.

Keywords: Anxiety, depression, sleep, pandemic, women

The COVID-19 pandemic, which originated in China and spread quickly throughout the world, was announced a "International Public Health Emergency" on January 30, 2020 [1], and a "Global Outbreak (pandemic)" on March 11, 2020 [2]. The COVID-19 virus, which has a very high transmis-

Received: March 7, 2025 **Accepted:** June 1, 2025 **Available Online:** June 4, 2025 **Published:** September 4, 2025

How to cite this article: Çatalgöl Ş, Akgün R. Associations between anxiety, depression, and sleep problems in women during the COVID-19 pandemic: a cross sectional study. Eur Res J. 2025;11(5):896-906. doi: 10.18621/eurj.1653062

Corresponding author: Şeyma Çatalgöl, RM, PhD., Phone: +90 276 221 21 21 ext. 3071, E-mail: seyma.catalgol@usak.edu.tr

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

sion rate, has affected the whole world and caused many people to die [3].

With the spread of the pandemic all over the world, countries have started to take their own measures. To control the spread of the pandemic; many more measures have been taken, such as social isolation, home quarantine, partial or complete lockdown [4].

Some measures have also been taken in Türkiye to prevent and control the spread of the epidemic. Some of these measures: such as closing schools and continuing education online [5], curfews for people under the age of 20 and above the age of 65 [6], travel restrictions, closure of shopping centers, and the public and private sector working from home [7]. With similar measures taken around the world, people began to spend more time at home, and accordingly, a major change in their lifestyle occurred [8, 9].

In addition, these changes have had negative effects on psychological health and sleep/wake patterns [10]. The new lifestyle that emerged in this process and the measures taken to control the pandemic caused an increase in anxiety and depression levels in individuals and led to sleep problems [11]. Maintaining homeostasis and a high standard of living in humans depends heavily on the process of sleep [12].

The incidence of disorders such as anxiety, depression, and stress is higher in individuals with sleep disorders. Especially in women, poor sleep is associated with some adverse health outcomes [13]. These are related to increases in comorbid conditions such as vasomotor symptoms, hormonal changes, age-related changes, and depression [14].

The likelihood of experiencing depression is also approximately twice as high in women as in men. This difference starts from adolescence and continues until the mid-50s. Therefore, women appear to be at greater risk of depression during their reproductive years [15].

During the COVID-19 pandemic, there was an increase in anxiety, depression, and stress levels due to the impact of lifestyle changes [3]. It is essential to recognize the psychological disorders and factors causing insomnia during the pandemic phase to prevent these problems. The purpose of this study was to evaluate the association between anxiety, depression, and sleep problems among women during the COVID-19 pandemic. To address the gaps mentioned above, this study has been conducted in line with the following research questions.

Research question 1: How has the COVID-19 pandemic affected women's anxiety, depression, and sleep quality levels?

Research question 2: Is there a statistically significant relationship between women's anxiety and depression levels and sleep problems?

Research question 3: How do demographic factors affect women's anxiety, depression, and sleep quality?

METHODS

Study Design

The data of this cross-sectional descriptive study was gathered from women who volunteered to participate in the research using the snowball sampling method across Türkiye between February 20, 2021 and April 9, 2021. Study questions were sent to women via Google Forms.

Inclusion Criteria

All women aged 18-65 years who gave electronic informed consent to participate in the study were included.

Exclusion Criteria

Women who had mental or sleep-related illnesses and were taking medication for these disorders and who gave logically incorrect answers to the study questions were excluded from the study.

Sample and Data Collection

The research universe is made up of 27974887 women between the ages of 18 and 65 in Türkiye [16]. The sample of the study was 686 women between the ages of 18 and 65 who voluntarily participated in the study from various parts of Türkiye. Introductory Features Information Form, Beck Anxiety Inventory (BAI), Beck Depression Inventory (BDI), and Pittsburgh Sleep Quality Index (PSQI) were used as measurement tools in the study.

The OpenEpi program Version 3 was used to determine the sample size. To determine the sample size, the following formula for descriptive research was used: Sample size (n)=[DEFF×Np(1-p)] / [(d2/Z21-α/2×(N-1)+p×(1-p)]. The minimum number of samples to be taken was calculated as 385 with 0.95 confidence and 0.05 margin of error with the sample

formula. To increase the efficiency of the study and the reliability of its results, 686 volunteer women were reached, and research data were collected. Answering every question was mandatory, and in the electronic form, unanswered questions could not be skipped. Therefore, there was no missing data. However, thirty participants were excluded from the analysis due to inconsistent responses.

Questionnaires

The Pittsburgh Sleep Quality Index, Beck Anxiety Inventory, and Beck Depression Inventory are the scales used in the study, and the information form created for the study was used to collect research data.

Information Form

The researchers developed 20 questions for the data collection form utilized in the study to assess the sociodemographic characteristics and sleep issues of women.

Pittsburgh Sleep Quality Index

The Pittsburgh Sleep Quality Index (PSQI) was developed by Buysse *et al*. [17] and adapted into Turkish by Ağargün *et al*. [18]. The total score is between 0 and 21. PSQI score >5 indicates poor sleep quality, while a PSQI score ≤ of 5 indicates good sleep quality. In this study, the coefficient of the scale in Cronbach's α was determined as 0.807.

Beck Anxiety Inventory

The Beck Anxiety Inventory (BAI), developed by Beck *et al*. [19], is a 21-item self-report inventory designed to assess the severity of anxiety symptoms using a 4-point Likert scale (0 = none/minimal, 1=mild, 2=moderate, 3=severe). The Turkish validity and reliability of the inventory were established by Ulusoy *et al*. [20]. Total scores range from 0 to 63, with 0-7 indicating minimal anxiety, 8-15 mild anxiety, 16-25 moderate anxiety, and 26-63 severe anxiety symptoms. In the current study, the Cronbach's alpha for the BAI was found to be 0.941.

Beck Depression Inventory

The Beck Depression Inventory (BDI) was developed by Beck *et al*. [21] to assess the risk and severity of depressive symptoms. The Turkish adaptation and validation were conducted by Hisli [22]. The scale's total score ranges from 0-63. In line with the corresponding score ranges, the scale includes 0-9 minimal, 10-16 mild mood disorders, 17-29 moderate depression, and between 30-63, which is considered severe depression. In this study, the Cronbach's α coefficient of the scale was determined as 0.922.

Statistical Analysis

The data obtained in the study were analyzed using the SPSS (Statistical Package for Social Sciences) for Windows 25.0 program. The assumption of normality was assessed statistically using the Kolmogorov-Smirnov test, and visually and descriptively through Q-Q plots and skewness-kurtosis values. The results of the Kolmogorov-Smirnov test indicated no significant deviation from normality. The difference between the two independent groups was assessed using an independent t-test. One-way analysis of Variance (ANOVA) was utilized to compare more than

Table 1. Socio-demographic characteristics of participants (n=686)

	n	%
Age		
18-25	444	64.7
26-35	137	20.0
36-45	52	7.6
46 and above	53	7.7
Marital Status		
Unmarried	480	70.0
Married	206	30.0
Education Level		
High college and lower	162	23.6
University	489	71.3
Post Graduate	35	5.1
Employment Status		
Student	171	24.9
Employed	281	41.0
Unemployed	234	34.1
Smoker		
Yes	189	27.6
No	497	72.4
Alcohol Consumption		
Yes	551	19.7
No	135	80.3

two independent groups, and the Bonferroni test was employed to identify the group from which the difference originated. To examine the relationships between the variables, Pearson correlation coefficients were calculated. This analysis was used to evaluate the linear associations between PSQI, BAI, and BDI scores.

To identify the factors affecting sleep quality among participants, a multiple linear regression analysis was conducted. In this analysis, the dependent variable was the PSQI score, and the independent variables were the BAI and BDI scores. This model was used to evaluate the predictive effects of anxiety and depression levels on sleep quality. $P<0.05$ was deemed to be statistically significant.

RESULTS

Information on the descriptive characteristics of the participants is provided in Table 1. In this study, 64.7% of the participants were between the ages of 18-25, 70.0% were single, 71.3% were university graduates,

Table 2. Sleep characteristics of participants during the pandemic period

	n	%
Have you had any changes in your waking hours during the pandemic?		
No change	280	40.8
I started waking up later than usual.	341	49.7
I'm waking up earlier than usual	65	9.5
Has there been a change in your sleep time compared to the pre-pandemic period?		
No change	287	41.8
I sleep less than before the pandemic	121	17.6
I sleep a lot compared to the pre-pandemic period.	278	40.5
How was your sleep quality during the pandemic period?		
Very good	40	5.8
Good	205	29.9
Middle	291	42.4
Bad	119	17.3
Very bad	31	4.5
How long are you exposed to sunlight during the day?		
None	103	15.0
1 hour or less	313	45.6
1-2 hours	147	21.4
2-3 hours	57	8.3
3-4 hours	32	4.7
4 hours or more	34	5.0
How much screens are you exposed to during the day compared to the pre-pandemic period?		
1 hour or less	64	9.3
1-2 hours	49	7.1
2-3 hours	84	12.2
3-4 hours	113	16.5
4 hours or more	376	54.8

Table 3. Sleep quality, anxiety and depression scores according to the characteristics of the participants

	PSQI	F[a]/t[b]	P value	BAI	F[a]/t[b]	P value	BDI	F[a]/t[b]	P value
Age									
18-25	6.61±2.86	2.606	0.051	15.80±13.13	1.304	0.272	16.76±12.26	9.452	**<0.001***
26-35	5.95±2.53			13.52±12.18			12.81±10.33		
36-45	5.98±3.26			14.18 ±11.59			11.27±9.44		
46 and above	6.02±3.14			14.22±13.16			10.62±9.65		
Marital Status									
Unmarried	6.59±2.91	2.892	**0.004***	15.73±13.13	1.997	**0.046***	16.41±12.27	4.606	**<0.001***
Married	5.90±2.71			13.60±12.05			11.97±9.71		
Education Level									
High college and lower	5.90±2.95	4.441	**0.012***	14.20±11.98	1.013	0.364	14.41±11.47	0.363	0.695
University	6.48±2.83			15.23±13.17			15.31±11.97		
Post graduate	7.31±2.72			17.45±12.22			14.83±9.54		
Employment Status									
Student	6.46±2.99	1.236	0.296	16.72±14.02	3.310	**0.037***	15.84±11.80	0.939	0.421
Employed	6.16±2.79			13.67±12.41			14.43±11.41		
Unemployed	6.58±2.85			15.62±12.85			15.24±12.07		
Smoker									
Yes	6.86±2.83	2.712	**0.007***	17.40±13.25	2.910	**0.004***	17.65±12.08	3.576	**<0.001***
No	6.20±2.86			14.22±12.60			14.10±11.46		
Alcohol Consumption									
Yes	7.50±2.74	-5.127	**<0.001***	19.71±13.63	-4.724	**<0.001***	18.03±12.60	-3.282	**0.001***
No	6.11±2.83			13.97±12.40			14.35±11.41		
Mean PSQI, BAI, BDI Scores	6.38±2.87			15.08±11.73			15.10±12.85		

Data are shown as mean±standard deviation. BAI=Beck Anxiety Inventory, BDI=Beck Depression Inventory, PSQI=Pittsburgh Sleep Quality Index.
[a]Tested by one-way ANOVA with post-hoc Bonferroni test. [b]Independent sample t-test *P<0.05

41.0% were employed, 27.6% smoked, and 19.7% drank alcohol.

The distribution of the participants according to their sleep characteristics during the pandemic period is presented in (Table 2). In our study, 49.7% of the participants reported waking up later than before the pandemic, 40.5% reported an increase in their sleep duration during the pandemic, 21.8% reported poor sleep quality, 45.6% reported being exposed to sunlight for one hour or less during the day compared to before the pandemic, and 54.8% reported being exposed to screens for four hours or more during the day. When examining the differences in sleep quality, anxiety, and depression scores (Table 3), significant differences were found between sleep quality and marital status, education, smoking, and alcohol consumption. The sleep quality of single individuals was worse than that of married individuals (P<0.05), and the sleep quality of postgraduate students was worse than that of individuals with a high school education or lower (P<0.05). Smokers and alcohol users had worse sleep quality compared to non-smokers and non-drinkers (P<0.05).

Significant differences were found in anxiety scores based on marital status, employment status, and tobacco and alcohol use. The anxiety scores of single individuals were higher than those of married individuals (P<0.05), and students had higher anxiety scores than working individuals (P<0.05). Smokers and alcohol users had higher anxiety scores compared to non-smokers and non-drinkers (P<0.05). Significant differences in depression scores were found based on age, marital status, smoking, and alcohol use. The depression scores of individuals in the 18-25 age group were higher than those in other age groups (P<0.05). Additionally, singles had higher depression scores compared to married individuals (P<0.05). Smokers and alcohol users also had higher depression scores compared to non-smokers and non-drinkers (P<0.05). The mean scores of the PSQI, BAI, and BDI scales used in the study were as follows: PSQI: 6.38±2.87, BAI: 15.08±11.73, and BDI: 15.10±12.85.

The relationship between sleep quality and anxiety and depression scores was found to be statistically significant (Table 4). As a result of the Bonferroni test, it was determined that the group with severe anxiety and depression had worse sleep quality than the other groups (P<0.05).

The results showing the relationship between the PSQI, BAI, and BDI scales used in the study are presented in (Table 5). The results show a moderate positive and significant relationship between sleep quality and anxiety (r=0.517; P<0.01), a moderate positive and significant relationship between sleep quality and depression (r=0.513; P<0.01), and a strong positive and significant relationship between anxiety and de-

Tablo 4. Sleep quality scores by anxiety and depression scores of the participants

	n	%	Mean±SD	F[a]	P value
Anxiety					
Minimal	250	36.44	4.73±2.12	78.274	**<0.001***
Mild	156	22.74	6.31±2.75		
Moderate	138	20.12	7.26±2.65		
Severe	142	20.70	8.52±2.58		
Total	686	100			
Depression					
Minimal	268	39.07	4.90±2.40	70.926	**<0.001***
Mild	159	23.18	6.40±2.58		
Moderate	173	25.22	7.40±2.60		
Severe	86	12.53	8.97±2.51		
Total	686	100			

SD=standard deviation. [a]Tested by one-way ANOVA with post-hoc Bonferroni test. *P<0.05

Table 5. Correlations between pittsburgh sleep quality index, beck anxiety inventory and beck depression inventory scores

Variables	PSQI	BAI	BDI
PSQI	-		
BAI	0.517**	-	
BDI	0.513**	0.647**	-

BAI=Beck Anxiety Inventory, BDI=Beck Depression Inventory, PSQI=Pittsburgh Sleep Quality Index
**P<0.01

pression (r=0.647; P<0.01).

The results of the multiple linear regression analysis show that anxiety (ß=0.071) and depression (ß=0.075) have a significant effect on sleep quality (P<0.05) (Table 6). With these effects, it is observed that 32% of the change in sleep quality is explained by anxiety and depression (Adjusted R^2=0.320). Additionally, the effect of anxiety on sleep quality was found to be higher than that of depression (Beta=0.318).

DISCUSSION

This study examined the anxiety, depression, and sleep problems experienced by women during the COVID-19 pandemic and the relationships between these problems. It was found that high anxiety and depression during the pandemic were significantly associated with poor sleep quality. Mandatory changes such as home quarantine, restriction of recreational activities, and transition to working from home led to significant changes in participants' mental health and sleep patterns [23]. In our study, various factors that may affect sleep quality were investigated and participants' responses to these factors were evaluated. It was determined that 49.7% of the participants woke up later during the pandemic compared to before the pandemic, 40.5% of them reported that their sleep duration increased compared to before the pandemic, and 45% were exposed to less than one hour of sunlight per day. Similarly, Raman and Coogan [24] reported that the participants in their study woke up later during the pandemic, and Rezaei and Grandner [25] reported that participants' sleep duration increased during COVID-19. Low exposure to sunlight can affect sleep quality by causing circadian rhythm disorders [26]. In our study, it was determined that 54.8% of the participants spent more than four hours in front of a screen. Majumdar *et al*. [27] also found that screen exposure increased during the pandemic. According to our study results, it is seen that women's sleep problems increased during the pandemic period.

In our study, it was found that 58.6% of women had poor sleep quality. Similar findings were obtained in other studies. Especially during the COVID-19 pandemic, women's PSQI scores were found to be high and their sleep quality was reported to be low [28-30]. It is seen that women have higher PSQI scores compared to men and sleep problems are more common [30-32].

Table 6. Multiple linear regression analysis of variables affecting participants' Pittsburgh Sleep Quality Index values

Dependent Variable	Independent Variables	ß	t	P value	VIF	Beta	F	Model (P)	Adjusted R^2
PSQI	**Constant**	4.181	27.416	**<0.001***					
	BAI	0.071	7.683	**<0.001***	1.721	0.318	162.298	**<0.001***	0.320
	BDI	0.075	7.446	**<0.001***	1.721	0.308			

BAI=Beck Anxiety Inventory, BDI=Beck Depression Inventory, PSQI=Pittsburgh Sleep Quality Index
Durbin Watson=1.869 *P<0.05

The mean BAI score of the participants was determined as 15.08±11.73. Smith *et al*. [34] found the mean BAI score of women as 14.1 ± 12.2. Massad *et al*. [35] reported that the mean BAI score of women in Jordan was 9.67. The high anxiety scores obtained in our study are not directly comparable because there are several differences between different populations. The mean BDI score of the participants in the current study was determined as 15.10 ± 12.85. Guadagni *et al*. [36] reported that the mean BDI score in women was 14.0 ± 9.9. Cellini *et al*. [10] reported that the mean BDI score in women students in their study in Italy was 9.93 ± 8.36. These differences may be due to differences in the age groups and sample size of the participants.

According to the findings obtained from our study, a significant relationship was found between sleep quality and some sociodemographic characteristics such as marital status, education level, smoking and alcohol use. It was observed that single participants had worse sleep quality than married participants. The results of the study by Deo *et al*. [37] are also consistent with these findings. It was observed that individuals who smoke and drink alcohol had higher sleep quality scores and that individuals in this group had poor sleep quality. These findings, in line with previous epidemiological studies, show that smoking and alcohol use have negative effects on sleep quality [38-40]. In our study, a significant relationship was found between the sociodemographic variables of age, marital status, smoking and alcohol use and the mean BDI scores. Similarly, a significant relationship was found between the sociodemographic variables of marital status, employment status, smoking and alcohol use and the mean BAI scores. When looking at age categories, it was found that women between the ages of 18-25 had higher depression scores than other age groups. Solomou and Constantinidou [41] stated that depression scores were higher in young individuals. It has been reported that young adults have poor sleep quality and high anxiety and depression risk [42]. In our study, it was determined that anxiety scores were higher in students. Wang *et al*. [43] found that students' anxiety levels were higher than other professional groups. Distance education in universities in Türkiye is thought to increase anxiety [44].

In our study, anxiety and depression scores were found to be higher in single participants than in married participants. Similarly, Gualano *et al*. [45] reported that anxiety and depression scores were higher in single participants in their study. Anxiety and depression scores were found to be higher in participants who smoked and drank alcohol. This finding is consistent with previous studies showing that individuals who smoked and drank alcohol are at higher risk for anxiety and depression [40, 46]. The pandemic period may pose potential risks for smoking and alcohol abuse and requires public health measures to protect vulnerable individuals. In this study, it was found that participants with severe depression and anxiety had higher sleep quality scores than other groups and that there was a significant relationship between sleep quality and anxiety and depression. In addition, it was found that the effect of anxiety on sleep quality was more pronounced than depression. Decreased sleep quality is strongly associated with poor mood [26]. Sleep problems are an important factor that can negatively affect psychological health [23].

Limitations

Among the limitations of the study, the sample size being restricted to a specific geographical area and the participants consisting solely of women can be mentioned. These limitations should be taken into account as they may restrict the generalizability of the findings. In future research, it is thought that studies conducted with large-scale and more heterogeneous samples, including individuals from different age groups as well as male participants, may provide more comprehensive results. Additionally, it would be beneficial to conduct studies that track the long-term psychological effects post-COVID-19 and evaluate the effectiveness of intervention strategies.

The study focused only on women, so the findings may not be generalizable to other populations. Future research should consider longitudinal studies and more diverse participant samples to understand these relationships in more depth.

CONCLUSION

This study reveals that the increased levels of anxiety and depression during the COVID-19 pandemic are significantly associated with poor sleep quality. The younger age group, students, singles, smokers, and al-

cohol users were at higher risk of increased anxiety, depression, and poor sleep quality during the pandemic.

According to our results, it is important to develop more targeted strategies for women, young people, and other vulnerable groups when creating health policies, to minimize the impacts of the pandemic. These strategies may include interventions such as ensuring easy access to psychological support services, increasing online therapy and support groups. Additionally, it is emphasized that early interventions and long-term monitoring for sleep disorders and psychological issues should be implemented. In this process, the adoption of community-based solutions through the collaboration of mental health professionals and healthcare institutions can alleviate the long-term psychological burden of the pandemic.

Ethical Statement

This study was approved by the Ethics Committee of Clinical Research, Faculty of Medicine, Uşak University (Date: 03.02.2021, Number: 35-35-11). We conducted this research in accordance with the guidelines in the Declaration of Helsinki. Informed consent was obtained from all participants.

Data Availability

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

Authors' Contribution

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

Conflict of interest

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

Financing

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

Generative Artificial Intelligence Statement

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

Editor's note

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

REFERENCES

1. World Health Organization. Novel Coronavirus (2019-nCoV) Situation Report–11. 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200131-sitrep-11-ncov.pdf?sfvrsn=de7c0f7_4. Accessed July 15, 2021.
2. World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19-11 March 2020. 2020. https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020. Accessed July 15, 2021.
3. Casagrande M, Favieri F, Tambelli R, Forte G. The enemy who sealed the world: effects quarantine due to the COVID-19 on sleep quality, anxiety, and psychological distress in the Italian population. Sleep Med. 2020;75:12-20. doi: 10.1016/j.sleep.2020.05.011.
4. Mayr V, Nußbaumer-Streit B, Gartlehner G. Quarantine alone or in combination with other public health measures to control COVID-19: A rapid review (review). Gesundheitswesen 2020;82(6):501-506. doi: 10.1055/a-1164-6611.
5. Turkish Council of Higher Education (CoHE). Press Release. Retrieved March 26. 2020. https://www.yok.gov.tr/Sayfalar/Haberler/2020/YKS%20Ertelenmesi%20Bas%C4%B1n%20A%C3%A7%C4%B1klamas%C4%B1.aspx. Accessed June 7, 2021.
6. Republic of Turkey, Ministry of Interior (2020, March 22). 2020. https://www.icisleri.gov.tr/65-yas-ve-ustu-ile-kronik-rahatsizligi-olanlara-sokaga-cikma-yasagi-ek-genelgesi. Accessed June 7, 2021.
7. Republic of Turkey, Ministry of Interior (2020, November 1). 2020. https://www.icisleri.gov.tr/koronavirus-ile-mucadele-kapsaminda-sokaga-cikma-kisitlamalari---yeni-kisitlama-ve-tedbirler-genelgeleri. Accessed June 7, 2021.
8. Wang J, Gong Y, Chen Z, *et al*. Sleep disturbances among Chinese residents during the Coronavirus Disease 2019 outbreak and associated factors. Sleep Med. 2020;74:199-203. doi: 10.1016/j.sleep.2020.08.002.

9. Aksoy A, Abiç A, Değirmenci F, Yılmaz DV. The relationship between quality of life and fear of Turkish individuals during the COVID-19 pandemic: A cross-sectional study. Arch. Psychiatr. Nurs. 2021;35(5):472-478. doi: 10.1016/j.apnu.2021.06.003.
10. Cellini N, Conte F, De Rosa O, et al. Changes in sleep timing and subjective sleep quality during the COVID-19 lockdown in Italy and Belgium: age, gender and working status as modulating factors. Sleep Med. 2021;77:112-119. doi: 10.1016/j.sleep.2020.11.027.
11. Duong CD. The impact of fear and anxiety of Covid-19 on life satisfaction: Psychological distress and sleep disturbance as mediators. Pers Individ Dif. 2021;178:110869. doi: 10.1016/j.paid.2021.110869.
12. Silva ESM, Ono BHVS, Souza JC. Sleep and immunity in times of COVID-19. Rev Assoc Med Bras (1992). 2020;66(Suppl 2):143-147. doi: 10.1590/1806-9282.66.S2.143.
13. Jackson ML, Sztendur EM, Diamond NT, Byles JE, Bruck D. Sleep difficulties and the development of depression and anxiety: a longitudinal study of young Australian women. Arch Womens Ment Health. 2014;17(3):189-198. doi: 10.1007/s00737-014-0417-8.
14. Pengo MF, Won CH, Bourjeily G. Sleep in women across the life span. Chest. 2018;154(1):196-206. doi: 10.1016/j.chest.2018.04.005.
15. Sassarini J. Depression in midlife women. Maturitas. 2016;94:149-154. doi: 10.1016/j.maturitas.2016.09.004.
16. Turkish Statistical Institute, İl, yaş grubu ve cinsiyete göre nüfus, 2007-2020, 2021. https://data.tuik.gov.tr/Bulten/Index?p=Adrese-Dayali-Nufus-Kayit-Sistemi-Sonuclari-2020-37210. Accessed: August 25, 2021.
17. Buysse DJ, Reynolds III CF, Monk TH, Berman SR, Kupfer DJ. The Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 1989;28(2):193-213. doi: 10.1016/0165-1781(89)90047-4.
18. Ağargün MY, Kara H, Anlar Ö. The validity and reliability of the Pittsburgh Sleep Quality Index. Turk J Psychiatry. 1996;7(2):107-115.
19. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol. 1988;56(6):893-897 doi: 10.1037//0022-006x.56.6.893.
20. Ulusoy M, Sahin NH, Erkmen H. Turkish version of the Beck Anxiety Inventory: psychometric properties. J Cogn Psychother. 1998;12(2):163-172.
21. Beck AT, Ward C, Mendelson M, Mock J, Erbaugh J. Beck depression inventory (BDI). Arch Gen Psychiatry. 1961;4(6):561-571.
22. Hisli N. Beck depresyon envanterinin universite ogrencileri icin gecerliligi, guvenilirligi. (A reliability and validity study of Beck Depression Inventory in a university student sample). J Psychol. 1989;7:3-13.
23. Fu W, Wang C, Zou L, et al. Psychological health, sleep quality, and coping styles to stress facing the COVID-19 in Wuhan, China. Transl Psychiatry. 2020;10(1):225. doi: 10.1038/s41398-020-00913-3.
24. Raman S, Coogan AN. Effects of societal-level COVID-19 mitigation measures on the timing and quality of sleep in Ireland. Sleep Med. 2022;91:179-184. doi: 10.1016/j.sleep.2021.02.024.
25. Rezaei N, Grandner MA. Changes in sleep duration, timing, and variability during the COVID-19 pandemic: large-scale Fitbit data from 6 major US cities. Sleep Health. 2021;7(3):303-313. doi: 10.1016/j.sleh.2021.02.008.
26. Targa AD, Benítez ID, Moncusí-Moix A, et al. Decrease in sleep quality during COVID-19 outbreak. Sleep Breath. 2021;25(2):1055-1061. doi: 10.1007/s11325-020-02202-1.
27. Majumdar P, Biswas A, Sahu S. COVID-19 pandemic and lockdown: cause of sleep disruption, depression, somatic pain, and increased screen exposure of office workers and students of India. Chronobiol Int. 2020;37(8):1191-1200. doi: 10.1080/07420528.2020.1786107.
28. Gaş S, Ekşi Özsoy H, Cesur Aydın K. The association between sleep quality, depression, anxiety and stress levels, and temporomandibular joint disorders among Turkish dental students during the COVID-19 pandemic. Cranio 2023;41(6):550-555. doi: 10.1080/08869634.2021.1883364.
29. Duran S, Erkin Ö. Psychologic distress and sleep quality among adults in Turkey during the COVID-19 pandemic. Prog Neuropsychopharmacol Biol Psychiatry. 2021;107:110254. doi: 10.1016/j.pnpbp.2021.110254.
30. Dogru BV, Kasar KS. The impact of the COVID-19 pandemic on sleep quality in the Turkish population: a cross-sectional study. P. R. Health Sci. 2022;41(2):82-88.
31. Hinz A, Glaesmer H, Brähler E, et al. Sleep quality in the general population: psychometric properties of the Pittsburgh Sleep Quality Index, derived from a German community sample of 9284 people. Sleep Med. 2017;30:57-63. doi: 10.1016/j.sleep.2016.03.008.
32. Idrissi AJ, Lamkaddem A, Benouajjit A, et al. Sleep quality and mental health in the context of COVID-19 pandemic and lockdown in Morocco. Sleep Med. 2020;74:248-253. doi: 10.1016/j.sleep.2020.07.045.
33. World Health Organization, 2020b. Gender and women's mental health. 2020. https://www.who.int/mental_health/prevention/genderwomen/en/. Accessed July 10, 2020.
34. Smith L, Jacob L, Yakkundi A, et al. Correlates of symptoms of anxiety and depression and mental wellbeing associated with COVID-19: a cross-sectional study of UK-based respondents. Psychiatry Res. 2020;291:113138. doi: 10.1016/j.psychres.2020.113138.
35. Massad I, Al-Taher R, Massad F, Al-Sabbagh MQ, Haddad M, Abufaraj M. The impact of the COVID-19 pandemic on mental health: early quarantine-related anxiety and its correlates among Jordanians. East Mediterr Health J. 2020;26(10):1165-1172. doi: 10.26719/emhj.20.115.
36. Guadagni V, Umilta' A, Iaria G. Sleep quality, empathy, and mood during the isolation period of the COVID-19 pandemic in the Canadian population: females and women suffered the most. Front Glob Womens Health. 2020;1:585938. doi: 10.3389/fgwh.2020.585938.
37. Deo PK, Budhathoki S, Raut J, Adhikari B, Shrestha J. Factors associated with perceived stress, anxiety, depression, insomnia during COVID-19 outbreak among nursing students. Int J Sci Res (IJSR). 2020;9(9):23-29. doi: 10.21275/SR20826170002.
38. Alqahtani SS, Banji D, Banji OJ. A survey assessing sleep efficiency among Saudis during COVID-19 home confinement

using the Pittsburgh sleep quality index: A call for health education. Saudi Pharm J. 2021;29(7):692-698. doi: 10.1016/j.jsps.2021.04.031.
39. AlAteeq D, Almokitib A, Mohideen M, AlBlowi N, Fayed A, Alshahrani SM. Stress, sleep, and use of sleep aids among physicians during the COVID-19 pandemic. J Nat Sci Med. 2021;4(2):197-204. doi: 10.4103/jnsm.jnsm_177_20.
40. Stanton R, To QG, Khalesi S, et al. Depression, anxiety and stress during COVID-19: associations with changes in physical activity, sleep, tobacco and alcohol use in Australian adults. Int J Environ Res Public Health. 2020;17(11):4065. doi: 10.3390/ijerph17114065.
41. Solomou I, Constantinidou F. Prevalence and predictors of anxiety and depression symptoms during the COVID-19 pandemic and compliance with precautionary measures: age and sex matter. Int J Environ Res Public Health. 2020;17(14):4924. doi: 10.3390/ijerph17144924.
42. Huang Y, Zhao N. Generalized anxiety disorder, depressive symptoms and sleep quality during COVID-19 outbreak in China: a web-based cross-sectional survey. Psychiatry Res. 2020;288:112954. doi: 10.1016/j.psychres.2020.112954.
43. Wang C, Pan R, Wan X, et al. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. Int J Environ Res Public Health. 2020;17(5):1729. doi: 10.3390/ijerph17051729.
44. Aslan H, Pekince H. Nursing students' views on the COVID−19 pandemic and their percieved stress levels. Perspect Psychiatr Care. 2021;57(2):695-701. doi: 10.1111/ppc.12597.
45. Gualano MR, Lo Moro G, Voglino G, Bert F, Siliquini R. Effects of Covid-19 lockdown on mental health and sleep disturbances in Italy. Int J Environ Res Public Health. 2020;17(13):4779. doi: 10.3390/ijerph17134779.
46. Clay JM, Parker MO. Alcohol use and misuse during the COVID-19 pandemic: a potential public health crisis? Lancet Public Health. 2020;5(5):e259. doi: 10.1016/S2468-2667(20)30088-8.

The European Research Journal 2025;11(5):907-914 Original Article

DOI: https://doi.org/10.18621/eurj.1641267 *Medical Physics and Deep Learning*

Pneumonia detection in chest X-ray images using convolutional neural networks

Çağdaş Şimşek[1], Suat Özkorukcuklu[1], Bora Işıldak[2]

[1]Deparment of Physics, Institute of Graduate Studies in Sciences, İstanbul University, İstanbul, Türkiye; [2]Deparment of Physics, Yıldız Technical University, Faculty of Art and Science, İstanbul Türkiye

ABSTRACT

Objectives: Pneumonia ranks among the infections and presents a considerable health threat, especially in certain age groups and developing countries. The accurate diagnosis of the disease and prompt identification are crucial for treatment purposes. This study aimed to develope a convolutional deep neural network model that can detect pneumonia using a sufficient number of chest X-ray images that have been verified with a "definite diagnosis" clinically.

Methods: This study uses a dataset that includes 1000 chest X-ray images from a variety of age groups taken as part of patient care at Koç University Faculty of Medicine Hospital Clinics. The dataset sample includes two sets of pictures called normal and pneumonia infected. Various preprocessing techniques were used on the obtained images, thus enabling the training and testing of our developed prediction model.

Results: We improved the accuracy of the model's decisions by applying image processing techniques, successfully achieving high levels of decision accuracy with our model We have elevated the precision of decision-making in our model to outstanding levels and achieved impressive F1 Score and AUC (Area Under the Curve) values (F1 Score: 0.94 and AUC Score: 0.98).

Conclusions: Our model was trained using X-ray images produced from the same devices of the same hospital and achieved very high prediction results, but using images produced from different countries, different hospitals and different devices, especially training and testing the model with much larger data sets, is a necessary need for this study and the model we developed to become more universal, and in this sense, there is a need to develop and expand the study.

Keywords: Pneumonia, convolutional neural network, chest X-ray, deep neural network, machine learning, deep learning, artificial intelligence

Studies have shown that pneumonia is a reason for hospital admissions in people of all ages and can be a serious health concern for the elderly population as it may result in severe illness and even death. A research investigation, into the number of deaths associated with pneumonia has revealed that national mortality statistics from 14 years of 19 european union countries were analysed and reported to be the most common cause of infection-related deaths, especially among the elderly and individuals with comorbidities. [1]. Children under 5 ages and individuals over 65 ages are, at risk of pneumonia related complications.

Received: February 18, 2025 **Accepted:** April 17, 2025 **Available Online:** July 23, 2025 **Published:** September 4, 2025

How to cite this article: Şimşek Ç, Özkorucuklu S, Işıldak B. Pneumonia detection in chest X-ray images using convolutional neural networks. Eur Res J. 2025;11(5):907-914. doi: 10.18621/eurj.1641267

Corresponding author: Çağdaş Şimşek, PhD., Phone: +90 212 455 57 00, E-mail: cagdas.simsek@cern.ch

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

In 2019 2.5 million deaths globally were linked to pneumonia [2].

Lung infection known as pneumonia can be caused by microorganisms like bacteria or viruses and is identified by respiratory symptoms and lung abnormalities on imaging tests such as infiltrations or consolidations in lung tissue or pleural effusion [3, 4].

The diverse range of infections across societies and age groups highlights a need to understand potential causes for effective treatment strategies. Identifying pathogens that cause lower respiratory tract infections, especially in childhood, is a challenging process because disease indicators may overlap with respiratory pathogens not related to infection, and existing routine tests may often lead to false-negative results or detect randomly transmitted pathogens [5].
Commonly seen in viral pneumonia are similar symptoms that can be observed by the doctor during a physical checkup to evaluate the patient's general well-being including respiratory difficulties and mental alertness. The doctor also checks the signs such as temperature and pulse rate along, with examining lung abnormalities by inspecting the rib cage and listening to the patients' breathing sounds. The results provide clues that suggest pneumonia may be present but to confirm the diagnosis definitively a chest X ray is necessary [6].

Alveolar infiltration seen on a chest X ray is described as the buildup of cells and fluid within the alveoli (air sacs). This pattern is frequently seen in pneumonia caused by bacteria such, as Streptococcus pneumoniae. On an X ray image alveolar infiltration appears as a dense opacity.
A common feature of pneumonia is the presence of boundaries usually limited to a specific part of the lung (referred to as lobar pneumonia).

Another characteristic is infiltration which occurs when inflammation and fluid buildup happen in the connective tissue between the air sacs that support lung structure. Known as the interstitium. This type of infiltration is often seen in cases of pneumonia or pneumonia caused by less common microorganisms like Mycoplasma pneumoniae. X ray images show infiltration as a broader and more spread-out pattern, with fine reticular or nodular features; however, this distinction may not always be clear cut. Some bacterial pneumonias like Staphylococcus aureus pneumonia can show infiltration while certain viral pneumonias may display alveolar infiltration as well. It's important to consider the findings along, with clinical observations and lab tests to determine the cause of pneumonia correctly [7].

Getting a chest X ray is a radiographic procedure used for screening purposes. Diagnosing lung conditions is commonly done using chest X ray imaging due to its cost effectiveness and ease of access without invasive procedures involved in the process of identification from the images relies heavily on skilled and seasoned doctors' knowledge and training " This challenge becomes particularly significant in regions, with limited healthcare resources where both the frequency and severity of these disorders are higher.

Pediatric pneumonia rates are much higher than the global average figures suggest. Automated diag-

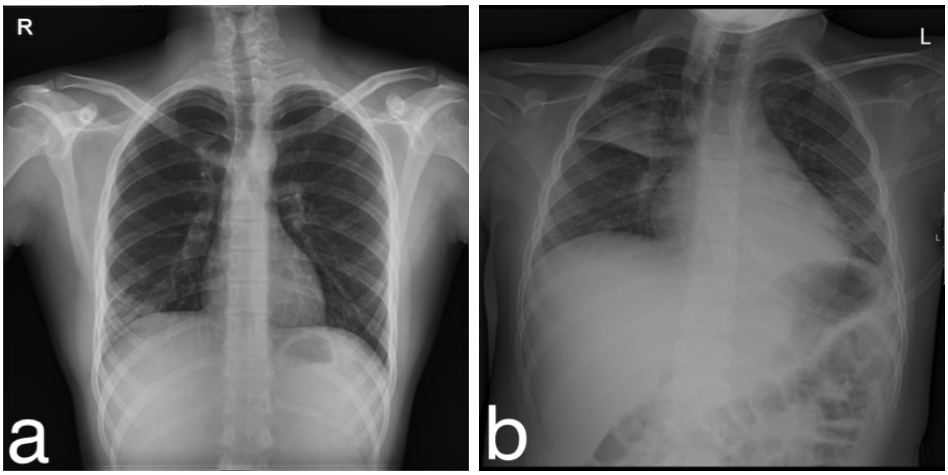

Fig. 1. Chest X-ray of a) normal patient and b) pneumonia infected patient.

nostic tools using chest X rays are seen as a way to improve radiologists' productivity and cut down on healthcare costs while speeding up the detection and treatment of pneumonia in kids.

METHODS

Experimental Dataset

This research study uses a dataset that includes 1000 chest X-ray images from a variety of age groups taken as part of patient care at Koç University Faculty of Medicine Hospital Clinics. The sample includes two sets of pictures called normal and pneumonia infected as shown in Fig. 1. Pneumonia is divided into bacterial and viral pneumonia. Qualified radiologists assigned labels during the data preparation process for patients, with a confirmed diagnosis based on chest X-ray images and advanced imaging methods as necessary. Backed up by laboratory tests and further validated through pathological examinations when needed and ultimately decided upon by the physicians.

The council's viewpoint was taken into consideration when organizing the imaging process and conducting laboratory procedures on samples from both healthy individuals and those with pneumonia infection for Convolutional Neural Network (CNN) training purposes later on; the images, in these sample subsets were then shuffled randomly for analysis as outlined in Table 1. Examining smaller portions, for analysis and evaluation.

Our approach to identifying pneumonia is based on the predictions we make. In order to guarantee the reliability and accuracy of the predictions made by our CNN model we have incorporated measures, into the image preprocessing stage. One, after the other algorithms are executed in sequence.

Table 1. Splitting of the dataset

Category	Training sample	Test sample
Normal images	190	60
Pneumonia infected images	565	185
Total images	755	245
Percentage of total images	75%	25%

Resize Pictures

The X-ray images of patients' chests can vary depending on factors like age and health status well as body type. In order to create a model, we meticulously adjusted all the images to ensure the lung, to image ratio is optimized. This manual adjustment led to images of sizes after cropping each one, as illustrated in Fig. 2. Therefore, we included the tactics mentioned below in our approach.

Preparing Images, for Analysis

After using the picture cropping method, we applied three processing steps to each image before including them in the training phase. Firstly, we adjusted all images to 512 × 512 size to match the CNN models

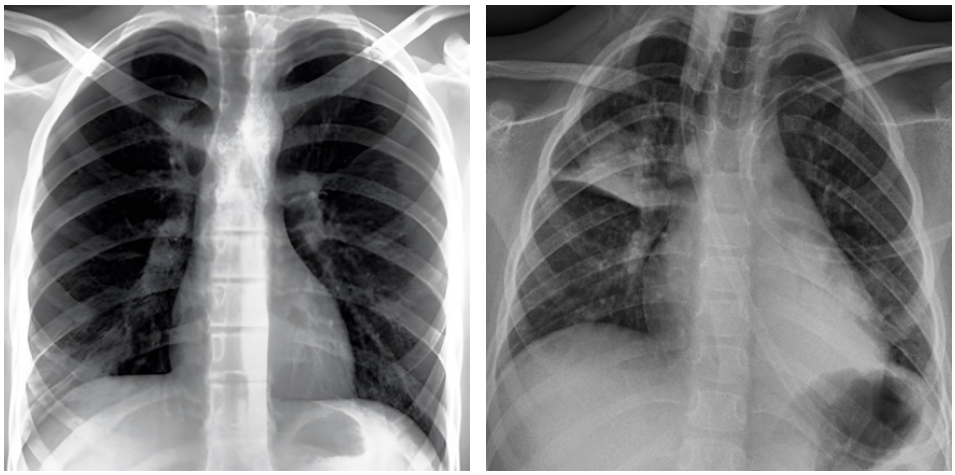

Fig. 2. Histogram equalized and cropped versions of the images shown in Fig. 1.

input requirements. We used bilinear interpolation to decrease the image sizes, for uniformity. In this step, we performed histogram equalization on the images to improve their overall quality and clarity.
Fig. 2 of the report is where you can see the effects of histogram equalization clearly shown in comparison to methods used in image processing tasks like normalizing pixel values within a specific range of 1 to 255 for better data scaling, in the third layer adjustment process.

Enhancing Data

We improved our sample by rotating it at an angle selected from a uniform range of [-10, 10] degrees. After the enhancement process was completed, we also doubled the size of the training sample set. Rotating images randomly boosts the model's robustness in accommodating the positioning of young patients and others. Severe illnesses can show images, on X-rays naturally.

Training at CNN

Many studies have shown results with various networks like ResNet18 [8], DenseNet121 [9], Xception [10], MobileNet [11]. However, these networks are designed for datasets. They have tens of millions of parameters for training A network with too many parameters trained with a small sample like ours often results in significant fluctuations, in loss values affecting accuracy values. Therefore the model cannot be considered.

Our study utilized a CNN design showcased in Fig. 3 for analysis purposes. During the training phase of the model development process running over 100 epochs with a batch size of 32. We opted for employing the square error (MSE) loss function paired with the Adam optimizer and maintained a learning rate set at 0.0001, during the training session.

Assessing Performance

The balanced datasets play a role in enhancing the predictive accuracy of both the Diagnostic and Diagnostic models in comparison to imbalanced datasets that require more complex neural network structures to effectively handle the data distribution disparities observed within the dataset where normal instances occur roughly four times less frequently than infected cases.

There seems to be a difference and inconsistency in the model predictions output observed here; however, if the model can maintain stability and accuracy even when trained under such conditions, at higher levels, it would give a significant edge in adjusting the model for various data sets broadly. The dataset comprises 93% of the information used for analysis purposes in this model's network configuration. When the number of layers and nodes increases, in the network model it tends to lead to a risk of overfitting, which can make the accuracy of the prediction algorithm more complex to achieve effectively.

Generating Grad-CAM Visualizations

In the realm of visual recognition tasks within

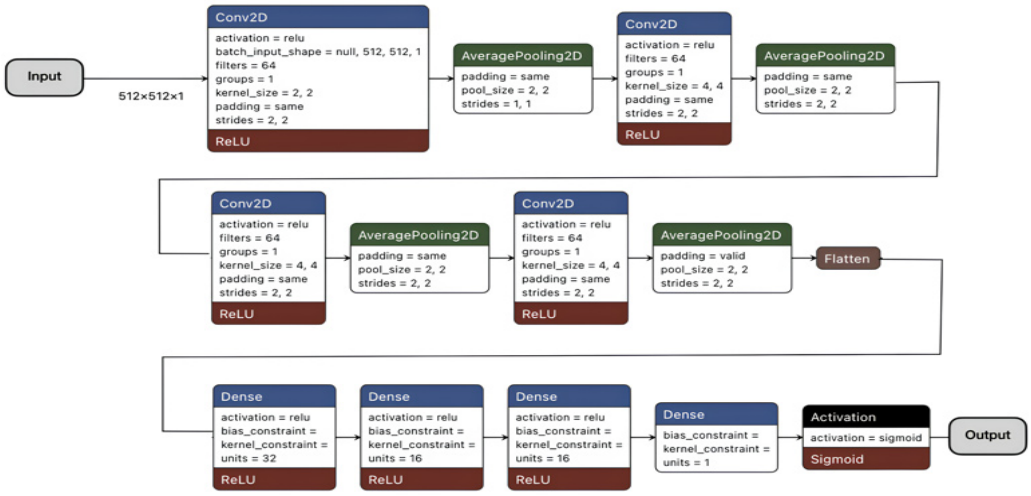

Fig. 3. Layer structure of Convolutional Neural networks with hyperparameters.

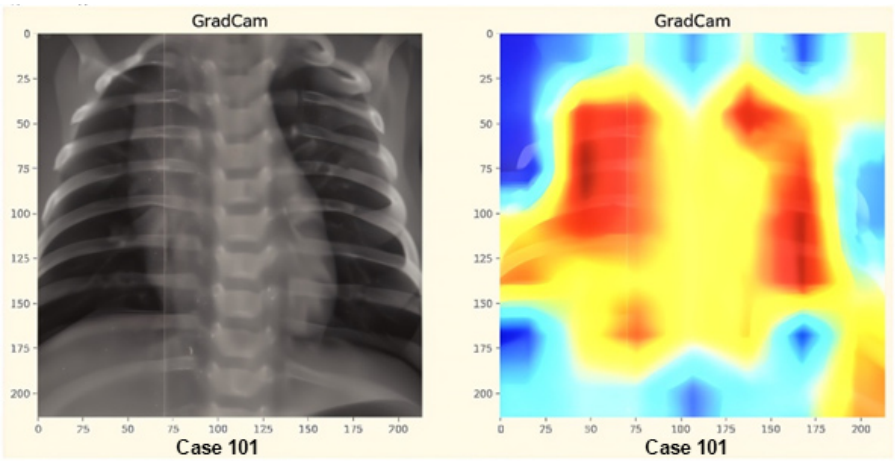

Fig. 4. When the model renders a judgment, the pixels it emphasizes are depicted in red.

deep learning models have achieved notable advancements over time. However understanding how these models make decisions proves to be a challenging endeavor. Shedding light on the characteristics that drive the model's decisions is crucial, for enhancing its reliability and pinpointing any shortcomings in this context.

Methods such as Grad CAM (Gradient weighted Class Activation Mapping) have emerged as a tool, for understanding how deep learning models work [12]. When applied to a model designed to detect pneumonia in chest X-rays, Grad CAM has demonstrated the model's ability to accurately identify pneumonia lesions [12]. We improved the clarity of the model's decisions for healthcare providers. Boosted its dependability, with Grad CAM technology that reveals which pixels the model focuses on most intensely in its decision-making process. This helps us better understand and improve the model's capabilities and weaknesses. We tested our model using a selected chest X-ray image and after it classified the image as either normal or abnormal.

The Grad CAM method was applied to analyze the image in Fig. 4 showcasing the areas where our model focused on while making decisions. The highlighted red regions indicate the key areas emphasized during the decision-making process.

Statistical Analysis

McNemar test was applied to evaluate the statistical significance of the classification performance of our model in pneumonia detection. As a result of the test using the number of false positives and false negatives in the test set, the test statistic was found to be 1.000 and the P-value was found to be 0.317.

This result shows that there is no statistically significant difference between the errors of the model ($P>0.05$). This shows that the performance of the model is balanced and not excessively biased towards a particular class. Therefore, the classification performance of the model is balanced and reliable.

RESULTS

In the sections for training and testing data sets the model accuracy results for both datasets are shown

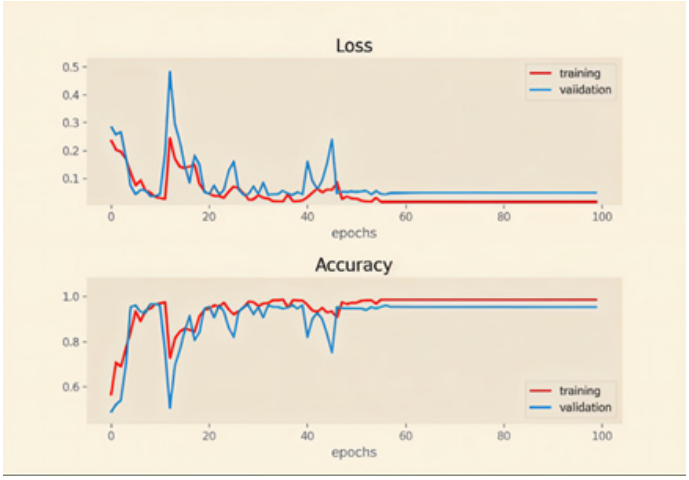

Fig. 5. Loss and accuracy curves for both training and validation sets during 100 epochs.

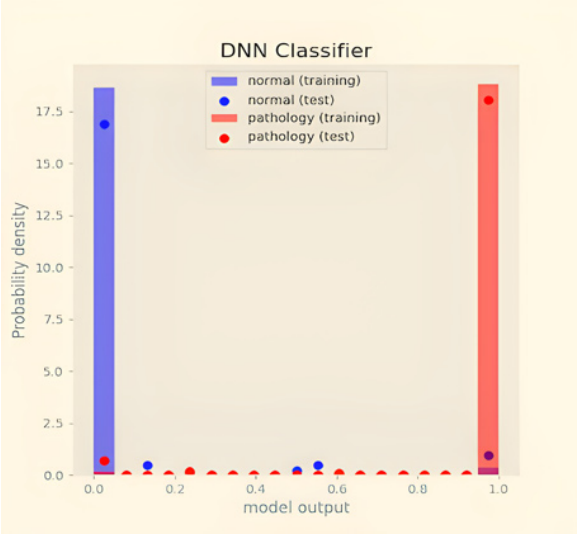

Fig. 6. CNN output distribution for normal images in the training sample (blue shaded), for the normal images in the test sample (blue dots), for pneumonia infected in the training sample (red shaded) and for pneumonia infected in the test sample (red dots).

below. The model attains 100% accuracy in the training set with its accuracy, in the test set being reached. The top part of Fig. 5 shows how the loss values change over training sessions. Fig. 6 shows the normalized output distributions of CNN for pneumonia-infected images, in both training and test sets are shown separately for each image type.

Table 2 displays the precision metrics for the model's performance assessment on the test sample using recall and F1 score values as well as the area under the Receiver Operating Characteristic (ROC) curve to quantify the obtained scores, in Fig. 7 left and right panels respectively.

DISCUSSION

Throughout the training phase of our model development, it's important to check for signs of overfitting to ensure performance. Taking a look at the distribution of model outputs for training and test data separately can help us assess whether the model is overfitting or not. Noticing a difference in the distributions of certain class labels can indicate potential issues with overfitting.

Inconsistencies between the training and test data could suggest that the model is overfitting the data for pneumonia detection. During the study period both pneumonia images from training and testing sets did not show any notable variances in results. The accuracy measure alone could have effectively assessed the performance of a model without overfitting had there existed a distribution of sample classes utilized in the training process. However, due to the absence of balance in our case, further metrics, like precision, recovery, F1 score, and AUC score have been incorporated for evaluation purposes.

The importance of the F-Score, especially in studies with dataset imbalances, is described in the study "A Survey of Evaluation Metrics for Classification Performance" [13]. The F-Score is important in areas such as medical diagnoses where the consequences of false negatives (for example, a cancer-positive case being described as cancer-negative) are crucial.

Limitations

In addition to all these, the limitations of the F1 score are discussed in one of the most important journals, the study titled "A Review of the F-Measure: Its History, Properties, Criticism, and Alternatives", which is one of the current studies published in ACM Computing Surveys (CSUR). For example, it is mentioned that it ignores the performance of the negative class while evaluating the performance of the positive class, and this situation can be problematic especially in multi-class classification problems [14].

CONCLUSION

In order to build a reliable and quantitative CNN model, we used an appropriate amount of data from chest X-ray images with a high-test accuracy of (93.3%) and

Table 2. Precision, recall, F1 score and AUC Score obtained for the test sample.

Pressicion	Recall	F1 Score	AUC Score
0.93	0.95	0.94	0.98

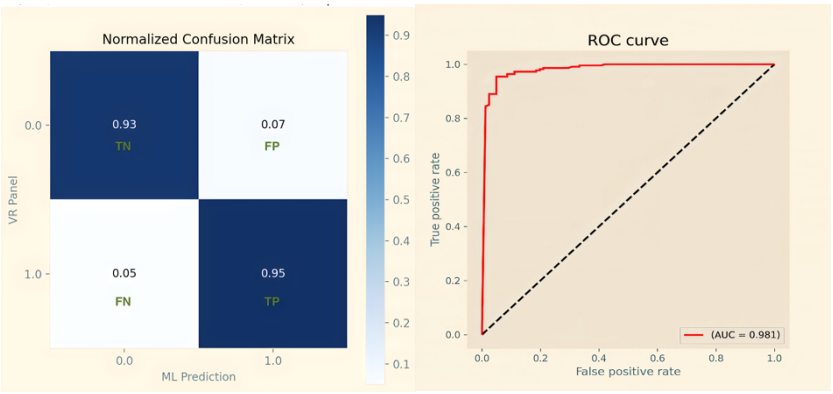

Fig. 7. Confusion Matrix obtained for the test sample. TN, TP, FN, FP stand for True Negative, True Positive, False Negative, False Positive respectively (left). ROC curve of the CNN model. The area under this curve gives the AUC score (right).

AUC score of (98.1%), using an adequate quantity of chest X-ray image data. The CNN model predictions are meant to expand the information available to clinicians during the decision-making process, especially in the case of pediatric triage. The purpose of presenting a prediction generated by artificial intelligence is to provide additional information for the radiologist in the decision-making process. The diagnosis has to be made by a specialist. On the other hand, there are places in the world where there are not enough doctors who can correctly identify pneumonia from chest X-ray images. A simple application on a smartphone can be used to gather better diagnostic information in these regions using the CNN model we have developed.

However, our model was trained using X-ray images produced from the same devices of the same hospital and achieved very high prediction results, but using images produced from different countries, different hospitals and different devices, especially training and testing the model with much larger data sets, is a necessary need for this study and the model we developed to become more universal, and in this sense, there is a need to develop and expand the study.

Ethical Statement

This study was approved by the Koç University Biomedical Research Ethics Committee (Decision no. 2025.300.IRB2.141, date: 30.06.2025).

Data Availability

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

Authors' Contribution

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

Conflict of interest

The authors disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The authors disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgement

All Chest X-ray images and the dataset used in this study were obtained from Koç University Faculty of Medicine Hospital.

I would also like to express my sincere gratitude to Prof. Dr. Mehmet Onur Demirkol for his valuable support and contributions in the clinical evaluation of medical images.

Generative Artificial Intelligence Statement

The author(s) declare that artificial intelligence tools were used in accordance with academic ethical standards during the preparation of this manuscript. Overleaf AI Assist tool was used in text editing oper-

ations such as checking spelling errors, editing words and adapting the reference source format. The all content of the study was produced by the author(s) in accor dance with scientific research methods and academic ethical principles.

Editor's note

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

REFERENCES

1. Marshall DC, Goodson RJ, Xu Y, et al. Trends in mortality from pneumonia in the Europe union: a temporal analysis of the European detailed mortality database between 2001 and 2014. Respir Res. 2018;19(1):81. doi: 10.1186/s12931-018-0781-4.
2. World Health Organization (WHO). Pneumonia in children. 2022. cited 2025 Feb 20. Available from: https://www.who.int/news-room/fact-sheets/detail/pneumonia
3. Mandell LA, Wunderink RG, Anzueto A, et al; Infectious Diseases Society of America; American Thoracic Society. Infectious Diseases Society of America/American Thoracic Society consensus guidelines on the management of community-acquired pneumonia in adults. Clin Infect Dis. 2007;44 Suppl 2(Suppl 2): S27-72. doi: 10.1086/511159.
4. Bartlett JG. Diagnostic tests for agents of community-acquired pneumonia. Clin Infect Dis. 2011;52 Suppl 4: S296-304. doi: 10.1093/cid/cir045.
5. Mick E, Tsitsiklis A, Kamm J, et al. Integrated host/microbe metagenomics enables accurate lower respiratory tract infection diagnosis in critically ill children. J Clin Invest. 2023 Apr 3;133(7):e165904. doi: 10.1172/JCI165904.
6. 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(7): e45-e67. doi: 10.1164/rccm.201908-1581ST.
7. File TM Jr, Ramirez JA. Community-Acquired Pneumonia. N Engl J Med. 2023;389(7):632-641. doi: 10.1056/NEJMcp2303286.
8. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770-778. doi: 10.1109/CVPR.2016.90.
9. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected Convolutional Networks. Densely Connected Convolutional Networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 2261-2269. doi: 10.1109/CVPR.2017.243.
10. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017, pp. 1800-1807. doi: 10.1109/CVPR.2017.195.
11. Howard AG, Zhu M, Chen B, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861. 2017 Apr 17.
12. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618-626. doi: 10.1109/ICCV.2017.74.
13. Hossin M, Sulaiman MN. A review on evaluation metrics for data classification evaluations. International Journal of Data Mining & Knowledge Management Process. 2015;5(2):1-11. doi: 10.5121/ijdkp.2015.5201.
14. Christen P, Hand DJ, Kirielle N. A Review of the F-Measure: Its History, Properties, Criticism, and Alternatives. ACM Comput. Surv. 2024;56(3):73. doi: 10.1145/3606367.

The European Research Journal 2025;11(5):915-921 Original Article

DOI: https://doi.org/10.18621/eurj.1692575 *Radiology*

AI-driven nodule detection in chest X-rays: Validation with radiologist-confirmed CT and X-ray findings

Yeliz Başar

Department of Radiology, Acıbadem Healthcare Group, İstanbul, Türkiye

ABSTRACT

Objectives: This study evaluates the performance of a commercially available deep learning tool in detecting chest nodules on X-rays, with findings validated by radiologist and confirmatory Computed Tomography (CT) scans.

Methods: In this retrospective analysis, the data of 299 consecutive patients who underwent both chest X-rays and CT scans within two weeks from June 2024 to December 2024 were analyzed. The performance of the deep learning tool was compared against the radiologist reports and CT scan reports, which were considered the gold standard. Performance parameters such as accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated and compared using McNemar's test.

Results: In 299 patients (43.8% female, mean age 58.78 years), CT scans showed nodules in 122 (40.8%) patients. Nodules were identified by the deep learning system in 53 patients, missing five cases identified by radiologist, but also identifying an additional 37 (30.3%) missed cases. The deep learning tool was 72% accurate (95% confidence interval: 0.66-0.77), higher than the 63% of the radiologist (P=0.02). Sensitivity was higher for artificial intelligence (37%) than for radiologist (10%; P<0.001) but lower in terms of specificity (95% vs. 100%; P=0.004).

Conclusions: The deep learning algorithm showed improved sensitivity and accuracy for the detection of pulmonary nodules but lower specificity, with reservations about false positives. The synergy of artificial intelligence and human experience has the potential to enhance the diagnosis of lung cancer. Further research is needed to validate these findings in diverse populations.

Keywords: Artificial intelligence, chest X-ray, chest nodüle, chest CT

Early detection of nodules in the lung is critical to early diagnosis and treatment of lung cancer, which remains one of the leading killer cancers worldwide [1]. Lung cancer is particularly insidious since it is often asymptomatic in its early stages, and therefore, early detection of nodules is crucial if survival is to be improved [2].

Traditional nodular detection with chest X-rays is greatly relies on radiologists' expertise. The mechanism, however, can be susceptible to various factors, including the tiredness of the radiologist, experience, as well as variation in human observation [3]. With the continued escalation in medical imaging exponentially, a greater demand for novel technologies able to increase efficiency and accuracy during diagnosis is felt. The amount of data generated by medical imaging

Received: May 20, 2025 **Accepted:** June 27, 2025 **Available Online:** July 16, 2025 **Published:** September 4, 2025

How to cite this article: Başar Y. AI-driven nodule detection in chest X-rays: Validation with radiologist-confirmed CT and X-ray findings. Eur Res J. 2025;11(5):915-921. doi: 10.18621/eurj.1692575

Corresponding author: Yeliz Başar, MD., Phone: +90 444 55 44, E-mail: yb772@hotmail.com

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

itself may be enormous and can be overwhelming even for the most skilled radiologists so that they may miss detection of a nodule [4]. To counter such setbacks, advancements in recent artificial intelligence have provided avenues for developing automated solutions capable of analyzing medical images with unparalleled precision [5].

Artificial intelligence (AI)-powered nodule detection software applies sophisticated deep learning algorithms to identify and classify chest X-ray abnormalities. These algorithms are trained on large datasets, allowing them to recognize patterns and features that may be pathognomonic [6]. By doing so, these programs have the potential to reduce the workload of radiologists, enhance diagnostic accuracy, and ultimately result in improved patient outcomes [7]. Yet, the clinical integration of these AI technologies requires strict validation compared to accepted diagnostic standards. It is important, as noted by a review performed by Najjar, that AI systems be thoroughly tested across various clinical environments to guarantee their robustness and performance in operational settings [8]. The effective deployment of AI for the detection of nodules can potentially revolutionize the discipline of lung cancer diagnosis by offering a supplement to human knowledge with the promise of quicker and more precise diagnoses [9].

The aim of this study is to assess the diagnostic accuracy and clinical utility of a commercially available deep learning tool in detecting pulmonary nodules on chest X-rays. The performance of the AI tool will be rigorously evaluated by comparing its findings with radiologist-confirmed detections on both chest X-rays and corresponding confirmatory Computed Tomography (CT) scans. This validation process seeks to establish the reliability of AI-driven nodule detection as a supportive tool in clinical practice, ensuring its potential to enhance early diagnosis and reduce oversight in chest radiography interpretation.

METHODS

Study Design and Sample

The review board approved this retrospective study and granted a waiver for informed consent regarding the collection, analysis, and presentation of anonymized medical data. This study follows the

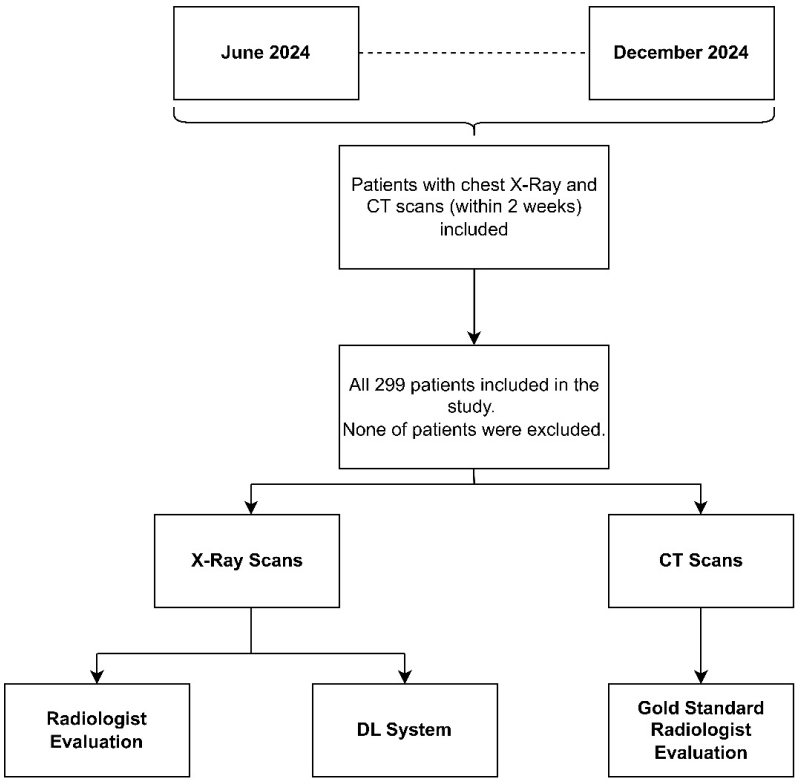

Fig. 1. Patient inclusion-exclusion and study design flowchart.

STROBE guideline [10].

We retrospectively reviewed consecutive patients who underwent chest X-ray scans and had CT scans within 2 weeks between June 2024 and December 2024. A total of 299 patients underwent chest X-ray scans and CT scans within 2 weeks. All 299 patients were included in the study. Patient inclusion-exclusion and study design were further described in Fig. 1.

Data Collection

All the scanning images were done in one center. The same X-ray (Siemens, Ysio, Munich, Germany) and CT (Siemens, NAEOTOM Alpha, Munich, Germany) units were used during the research. European Medical Device Regulation (MDR) CE-marked commercially available deep learning tool hChestXR (v1.0, Hevi AI, Istanbul, Turkey) was used to analyze nodules in chest X-rays. hChestXR tool was directly installed in the hospital's Picture Archiving and Communication System (PACS). hChestXR directly analyzed images from hospital PACS. Results were noted. A single radiologist with 20 years of experience in chest imaging interpreted all the X-ray and CT scan images. CT scans were used as a gold standard. None of the images used were used in the deep learning model development.

Statistical Analysis

All the analysis was done by R statistical software (Austria, R Core Team, version 4.1.0). Statistically significant was taken as a 0.95 confidence interval. Descriptive analysis was done on patient age and sex. The deep learning system and radiologist chest X-ray reading evaluation were measured with a confusion matrix. Accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated and compared with the McNemar test.

RESULTS

From all patients 131 (43.8%) of them were female and 168 (56.2%) of them were male. Mean age was 58.78 (±18.9). Further information was given in Fig. 2. CT examinations revealed 122 (40.8%) patients had chest nodules. Radiologist found chest nodules in 12 (12/122, 9.8%) patients from chest X-rays of those patients without any false positive cases. deep learning system able to detect chest nodules in 53 patients. The deep learning system missed five (5/12, 41.6%) cases that radiologist found. Of 53 patients that deep learning assessed nodule, 8 (8/53, 15%) of them were false

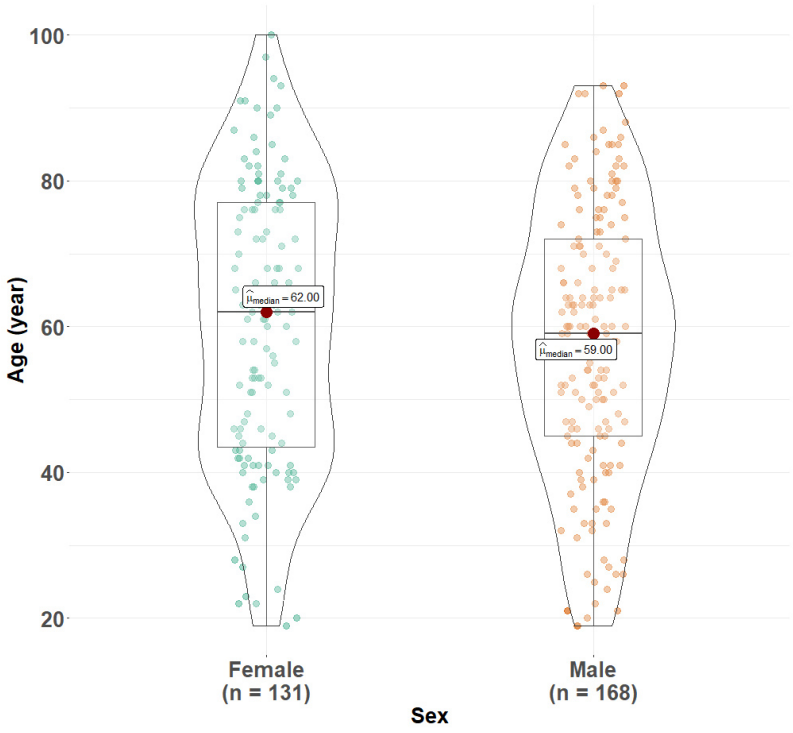

Fig. 2. Age and sex distribution of the study.

Table 1. Diagnostic metrics for pathologies

	Accuracy	Sensitivity	Specificity	PPV	NPV
Radiologist X-ray assessment	0.63 (95% CI: 0.57-0.69)	0.10 (95% CI: 0.05-0.17)	1.00 (95% CI: 0.98-1.00)	1.00 (95% CI: 0.74-1.00)	0.62 (95% CI: 0.56-0.67)
AI	0.72 (95% CI: 0.66-0.77)	0.37 (95% CI: 0.28-0.46)	0.95 (95% CI: 0.91-0.98)	0.85 (95% CI: 0.72-0.93)	0.69 (95% CI: 0.63-0.74)
P value	**0.02**	**<0.001**	**0.004**	**0.005**	**<0.001**

PPV=Positive predictive value, NPV=Negative predictive value, AI=Artificial intelligence, CI=Confidence Interval.

positives. Deep learning system able to detect extra 38 (38/122, 31.1%) different patients that radiologist unable to detect nodule in X-ray. The deep learning system achieved an accuracy of 0.72 (95% confidence interval [CI]: 0.66-0.77), which was statistically significantly higher than the radiologist's accuracy of 0.63 (95% CI: 0.57-0.69; P=0.02). Sensitivity was statistically significantly higher for deep learning, reaching 0.37 (95% CI: 0.28-0.46), compared to the radiologist's sensitivity of 0.10 (95% CI: 0.05-0.17; P<0.001). Specificity was statistically significantly lower for deep learning at 0.95 (95% CI: 0.91-0.98), radiologist demonstrated specificity of 1.00 (95% CI: 0.98-1.00; P=0.004). PPV was statistically significantly higher for radiologist with 1.00 (95% CI: 0.74-1.00) and deep learning with 0.85 (95% CI: 0.72-0.9; (P=0.005), while NPV was statistically significantly higher for DL (0.69, 95% CI: 0.63-0.74) than radiologist (0.62, 95% CI: 0.56-0.67) (P<0.001). Further information can be found in Table 1.

DISCUSSION

In the study research, the performance of a commercially deployed deep learning system in the detection of chest nodules on X-rays evaluated with all the results validated by radiologist through X-rays and confirmatory CTs. Study highlights the potential use of deep learning-based detection of nodules in chest X-rays, with its potential for enhancing diagnostic accuracy in the detection of pulmonary nodules. The deep learning model achieved a statistically significant accuracy of 72%, compared to 63% by the radiologist, and achieved significantly higher sensitivity, which is the ratio of correctly detected nodules by the model. Specifically, better accuracy for hard-to-detect nodules on X-ray. But the specificity of the deep learning was lower than that of the radiologist, which means a higher rate of false positives. The results of the current study support previous studies highlighting the role of AI in increasing the diagnostic correctness of pulmonary nodule detection from chest X-rays.

For instance, a systematic review by Ramli *et al*. [11] showed that AI algorithms were capable of generating wide ranges of sensitivity in identifying lung nodules based on the assessment strategy. Gold standard modality was the key determinant whether it is

CT or X-ray based [11]. In our study demonstrated comparable sensitivity with CT scans, gold standard. Where deep learning tool had a statistically significantly lower specificity of 0.95 while being compared to the specificity of the radiologist at 1.00. Deep learning tool did equally better for literature [12-15]. Additionally, the deep learning ability to detect more nodules not identified by the radiologist (30.3%) is in agreement with prior studies by Ardila *et al*. [16] on CTs, which indicated deep learning's utility in assisting radiologists by detecting difficult-to-detect abnormalities. Ahmad *et al*. [17] compared a number of studies on the use of AI in radiology and found that while AI algorithms were extremely sensitive in the detection of lung nodules, specificity was often less than ideal. Such a difference can lead to more false positives, and this could result in more biopsies or further imaging workup. The authors concluded that while AI could prove to be a helpful ally to traditional radiological practice, it should never replace the clinical expertise of the radiologist [17]. Together, these reports underscore the adjunctive role of AI in radiology as suggesting the potential of AI to add to nodule detection, but with caution in the implementation to avoid the dangers of false positives. The increasing burden of the interpretation of chest X-rays is a significant issue for radiologists, particularly against the backdrop of today's exploding volume of medical images and the complexity of the identification of pulmonary nodules [18]. Regular radiological assessments are usually marred by human errors, suffering from the effects of fatigue and unevenness in expertise, which consequently can generate diagnostic and tardy treatment errors [19]. Integration of AI into this process has proven to be a good method for enhancing accuracy and the effectiveness of the diagnostic process [20]. While AI has the ability to alleviate some burden in relation to chest X-ray interpretation, its application must be pursued with care in the attempt to restrict the opportunities of inappropriate intervention due to false positives [11].

Limitations

Despite the encouraging results of this study, certain limitations should be noted. First, the retrospective design might lead to bias. Second, the study was done at a single center, which could limit the ability to generalize the findings to larger groups or other clinical settings. Third, use of a single radiologist to interpret all images may introduce some level of subjectivity, potentially skewing the consistency of the readings. Finally, the deep learning tool was evaluated for its performance with a small dataset, and further research must be done to establish its effectiveness on heterogeneous populations and imaging scenarios to show its resilience in clinical practice.

CONCLUSION

In conclusion, this research suggests the potential for deep learning -augmented nodule detection software to improve chest X-ray diagnostic precision in pulmonary nodules. The lower specificity does present the issue of false positives, which might cause patient anxiety and unnecessary interventions. Future studies need to replicate these findings in different populations and settings. Ultimately, the synergy between AI and human expertise might enhance lung cancer diagnostics, and thereby management.

Ethical Statement

All procedures conducted in studies involving human participants complied with the ethical standards of the institutional and/or national research committee, as well as the 1964 Helsinki Declaration and its subsequent amendments or equivalent ethical guidelines. The local ethics committee (Acıbadem Mehmet Ali Aydınlar University Medical Research Evaluation Board - ATADEK) approved this retrospective study and waived the requirement for informed consent for the retrospective analysis of anonymized medical data (Decision number: 2025-07/52 and date: 08.05.2025).

Data Availability

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

Authors' Contribution

Study Conception: YB; Study Design: YB; Supervision: YB; Funding: YB; Materials: YB; Data Collection and/or Processing: YB; Statistical Analysis

and/or Data Interpretation: YB; Literature Review: YB; Manuscript Preparation: YB; and Critical Review: YB.

Conflict of interest

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

Financing

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

Generative Artificial Intelligence Statement

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

Generative Artificial Intelligence Statement

The author(s) declare that artificial intelligence tools were used in accordance with academic ethical standards during the preparation of this manuscript. European Medical Device Regulation (MDR) CE-marked commercially available deep learning tool hChestXR (v1.0, Hevi AI, Istanbul, Turkey) was used to analyze nodules in chest X-rays. hChestXR tool was directly installed in the hospital's Picture Archiving and Communication System (PACS). hChestXR directly analyzed images from hospital PACS. The all content of the study was produced by the author(s) in accordance with scientific research methods and academic ethical principles.

Editor's note

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

REFERENCES

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024;74(1):12-49. doi: 10.3322/caac.21820.
2. Hendriks LEL, Remon J, Faivre-Finn C, et al. Non-small-cell lung cancer. Nat Rev Dis Primers. 2024;10(1):71. doi: 10.1038/s41572-024-00551-9.
3. Cheng PM, Montagnon E, Yamashita R, et al. Deep Learning: An Update for Radiologists. Radiographics. 2021;41(5):1427-1445. doi: 10.1148/rg.2021200210.
4. Huang S, Yang J, Shen N, Xu Q, Zhao Q. Artificial intelligence in lung cancer diagnosis and prognosis: Current application and future perspective. Semin Cancer Biol. 2023;89:30-37. doi: 10.1016/j.semcancer.2023.01.006.
5. Esteva A, Robicquet A, Ramsundar B, et al. A guide to deep learning in healthcare. Nat Med. 2019;25(1):24-29. doi: 10.1038/s41591-018-0316-z.
6. Lakhani P, Sundaram B. Deep Learning at Chest Radiography: Automated Classification of Pulmonary Tuberculosis by Using Convolutional Neural Networks. Radiology. 2017;284(2):574-582. doi: 10.1148/radiol.2017162326.
7. Bramley R, Broadhurst P, Sharman A, et al. Prevalence of AI findings on chest X-ray in patients with lung cancer: A cross-sectional cohort study. medRxiv.2024.12.20.24319410. [Preprint]. doi: 10.1101/2024.12.20.24319410.
8. Najjar R. Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. Diagnostics (Basel). 2023;13(17):2760. doi: 10.3390/diagnostics13172760.
9. Anderson PG, Tarder-Stoll H, Alpaslan M, et al. Deep learning improves physician accuracy in the comprehensive detection of abnormalities on chest X-rays. Sci Rep. 2024;14(1):25151. doi: 10.1038/s41598-024-76608-2.
10. 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(4):344-3449. doi: 10.1016/j.jclinepi.2007.11.008.
11. Megat Ramli PN, Aizuddin AN, Ahmad N, Abdul Hamid Z, Ismail KI. A Systematic Review: The Role of Artificial Intelligence in Lung Cancer Screening in Detecting Lung Nodules on Chest X-Rays. Diagnostics (Basel). 2025;15(3):246. doi: 10.3390/diagnostics15030246.
12. Peters AA, Decasper A, Munz J, et al. Performance of an AI based CAD system in solid lung nodule detection on chest phantom radiographs compared to radiology residents and fellow radiologists. J Thorac Dis. 2021;13(5):2728-2737. doi: 10.21037/jtd-20-3522.
13. Govindarajan A, Govindarajan A, Tanamala S, et al. Role of an Automated Deep Learning Algorithm for Reliable Screening of Abnormality in Chest Radiographs: A Prospective Multicenter Quality Improvement Study. Diagnostics (Basel). 2022;12(11):2724. doi: 10.3390/diagnostics12112724.
14. Ahn JS, Ebrahimian S, McDermott S, et al. Association of Artificial Intelligence-Aided Chest Radiograph Interpretation With Reader Performance and Efficiency. JAMA Netw Open. 2022;5(8):e2229289. doi: 10.1001/jamanetworkopen.2022.29289.
15. Koo YH, Shin KE, Park JS, Lee JW, Byun S, Lee H. Extravalidation and reproducibility results of a commercial deep learning-based automatic detection algorithm for pulmonary nodules on chest radiographs at tertiary hospital. J Med Imaging Radiat

Oncol. 2021;65(1):15-22. doi: 10.1111/1754-9485.13105.
16. Ardila D, Kiraly AP, Bharadwaj S, et al. End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography. Nat Med. 2019;25(6):954-961. doi: 10.1038/s41591-019-0447-x.
17. Ahmad HK, Milne MR, Buchlak QD, et al. Machine Learning Augmented Interpretation of Chest X-rays: A Systematic Review. Diagnostics (Basel). 2023;13(4):743. doi: 10.3390/diagnostics13040743.
18. Delrue L, Gosselin R, Ilsen B, Van Landeghem A, de Mey J, Duyck P. Difficulties in the Interpretation of Chest Radiography. In: Coche E, Ghaye B, de Mey J, Duyck P. Editors. Comparative Interpretation of CT and Standard Radiography of the Chest. Springer, Berlin, Heidelberg. 2011: pp. 27-49.
19. Mazzone PJ, Lam L. Evaluating the Patient With a Pulmonary Nodule: A Review. JAMA. 2022;327(3):264-273. doi: 10.1001/jama.2021.24287.
20. Nam JG, Hwang EJ, Kim J, et al. AI Improves Nodule Detection on Chest Radiographs in a Health Screening Population: A Randomized Controlled Trial. Radiology. 2023;307(2):e221894. doi: 10.1148/radiol.221894.

The European Research Journal 2025;11(5):922-933 Original Article

DOI: https://doi.org/10.18621/eurj.1684936 *Public Health and Artificial Intelligence*

Assessing AI-based chatbots accuracy in caloric estimation: A focus on traditional Turkish foods

Hüsna Kaya Kaçar, Buse Sarıkaya

Department of Nutrition and Dietetics, Faculty of Health Sciences, Amasya University, Amasya, Türkiye

ABSTRACT

Objectives: To evaluate the accuracy of three widely used AI-based chatbots - ChatGPT 4.0, Microsoft Copilot, and Gemini - in estimating the caloric values of traditional Turkish foods.

Methods: The accuracy of caloric information provided by the chatbots for 71 traditional Turkish foods selected from the Türkomp National Food Composition Database was assessed. Each chatbot was queried in Turkish using the standardised prompt: "What is the caloric content of [food name] per 100 grams?" Responses were recorded and, when necessary, converted to a per-100-gram basis to ensure consistency. Accuracy percentages were calculated by comparing chatbot responses to Türkomp reference values, and extreme deviations were adjusted to 0%. Mean accuracy scores and distribution across predefined accuracy intervals were analysed for each chatbot. Statistical analysis was conducted to determine the mean accuracy percentages and identify differences among the chatbots.

Results: ChatGPT 4.0 achieved the highest mean accuracy (81.62%±20.6%), followed closely by Microsoft Copilot (81.23%±20.7%), while Gemini demonstrated lower accuracy (70.99%±30.2) ($P<0.05$). A one-way ANOVA showed a statistically significant difference in the mean accuracy percentages among the chatbots ($F(2, 210)=4.39$, $P=0.0136$). Foods such as kefir, tahini halva, and walnut baklava were estimated with over 90% accuracy by all three chatbots, suggesting strengths in their training datasets and the relatively simple or standardised nutrient composition of these foods. However, significant discrepancies in caloric estimations were observed across the chatbots, likely due to differences in algorithms and database integrations.

Conclusions: The findings suggest that AI-based chatbots have the potential to serve as culturally relevant tools for dietary assessment. However, the results also emphasise the need for careful use and further development. While ChatGPT 4.0 and Microsoft Copilot performed better than Gemini, the study shows the need for improved algorithms and expanded training datasets to enhance the accuracy and reliability of chatbots in nutritional evaluation. This study contributes to the growing body of research on AI applications in dietetics and public health; however, addressing these limitations is crucial to ensure their practical utility. Optimising chatbot design for real-world use will require interdisciplinary collaboration among AI developers, nutritional scientists, and healthcare professionals.

Keywords: Artificial intelligence, caloric estimation, chatbot accuracy, dietary assessment, nutrition technology, traditional Turkish foods

Received: April 27, 2025 **Accepted:** August 14, 2025 **Available Online:** August 18, 2025 **Published:** September 4, 2025

How to cite this article: Kaya Kaçar H, Sarıkaya B. Assessing AI-based chatbots accuracy in caloric estimation: A focus on traditional Turkish foods. Eur Res J. 2025;11(5):922-933. doi: 10.18621/eurj.1684936

Corresponding author: Hüsna Kaya Kaçar, PhD., Assist. Prof., Phone: +90 232 329 35 35, E-mail: husna.kacar@amasya.edu.tr

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

Artificial intelligence (AI)-based chatbots have rapidly integrated into various domains, offering tools for information access and decision support [1-3]. In the field of health and nutrition, these tools are increasingly being employed to provide dietary recommendations and nutritional information including the caloric content of foods [4, 5]. Their growing prevalence in everyday life highlights their potential to transform health and nutrition practices [6-8]. However, despite their convenience, questions about the accuracy and reliability of the nutritional data they provide, particularly in culturally specific contexts, persist as a significant concern [9-11].

Previous research has identified significant discrepancies in chatbot-generated dietary information, emphasising the need for systematic evaluations of their accuracy [12]. There is a growing body of literature recognised that chatbots often provide inconsistent nutritional advice, especially when trained on diverse, unverified data sources, which can compromise their reliability [13, 14]. These inconsistencies can arise from the diverse range of sources that chatbots rely on for nutritional data, including both commercial and open-access databases [15, 16]. Their responses often exhibit notable limitations and inconsistencies while large language models (LLMs) demonstrate potential in providing nutritional information [17]. These challenges may appear from factors such as the quality and scope of their training data, the architecture of the models, and the strategies used to frame user prompts [10, 11, 18]. Therefore, the precision of energy content in food data is a critical aspect that needs closer examination in the context of AI-based dietary tools.

Recent advances in LLMs have transformed natural language processing (NLP) [19]. This enables powerful general-purpose tools such as ChatGPT, Microsoft Copilot, and Gemini to generate fluent, contextually appropriate responses across a wide range of domains [20, 21]. While general performance has been well studied [22-24], there remains a critical gap in understanding how LLMs perform in domain-specific factual estimation tasks, such as calculating the caloric content of traditional food items. These tasks demand not only linguistic fluency but also factual grounding in nutritional science and region-specific culinary knowledge. Evaluating the ability of LLMs to accurately respond to structured, factual queries in specialised domains is an unfold challenge for NLP [25-27]. Previous work has largely focused on medical question answering [28, 29], legal reasoning [30-32], and financial forecasting [33, 34]; however, the nutritional domain remains underexplored, especially concerning non-Western food systems. From an NLP perspective, this evaluation contributes to ongoing discussions around trustworthy AI, model grounding, and bias in domain transfer. By systematically comparing outputs from three widely used LLMs, this research provides evidence on the models' factual accuracy, revealing the limits of their current training data and inference strategies when confronted with culturally embedded, non-standardised inputs. Findings of this study are critical for developing stronger, domain-aware models and for informing the responsible use of AI in health-related decision-making.

From a health perspective, accurate caloric information is essential for individuals managing their weight, following specific dietary plans, or making informed food choices [35, 36]. Traditional Turkish cuisine, renowned for its rich cultural heritage and diverse range of dishes, presents distinct challenges for accurate caloric estimation due to its unique ingredients, varied preparation methods, and portion size variability [37, 38]. These dishes often lack standardised portion sizes and are influenced by home-cooking practices, such as the use of oils, spices, and preparation methods like frying, boiling, or baking, all of which significantly affect energy content and nutrient composition [39, 40]. These complexities often lead to nutritional profiles that are not fully captured by generalised food databases, resulting in difficulties for AI systems trained on such data [11, 41, 42]. AI-based tools and chatbots, typically trained on Western-centric or generalised food databases, struggle to interpret such culturally specific foods with accuracy [43, 44]. Additionally, these systems often lack the contextual understanding necessary to estimate portion sizes visually or account for preparation techniques that impact the final caloric value [45-48]. As reliance on AI chatbots for nutrition-related queries continues to grow [11, 13, 16], assessing their performance in estimating the caloric content of culturally specific foods, such as those from Turkish cuisine, becomes increasingly important.

This study aims to address these gaps by assessing the calorie estimation accuracy of three widely used

AI chatbots-ChatGPT 4.0, Microsoft Copilot, and Gemini-specifically for traditional Turkish foods [49-51]. These chatbots were selected based on their significant market presence, accessibility to consumers, and underlying technological distinctions. ChatGPT 4.0, developed by OpenAI, is known for its natural language understanding and multimodal capabilities, making it one of the most widely adopted AI tools for general-purpose queries, including nutrition-related advice [21, 52]. Microsoft Copilot incorporates OpenAI's models within Microsoft's platform, providing users with easy access to AI functions through commonly used applications such as Bing and Microsoft Office [53, 54]. Gemini, previously known as Bard AI, developed by Google DeepMind, represents a different lineage of LLMs and is deeply integrated into Google's search and mobile infrastructure, positioning it as a major competitor in the generative AI space [55]. By comparing the chatbots' responses against the reference data provided by the Türkomp National Food Composition Database [56], this research seeks to interpret their performance and reliability. This study addresses a foundational question: To what extent can current LLMs accurately estimate caloric values for culturally specific foods? The findings of this study will not only inform the development of more precise AI-driven nutrition tools but also contribute to the broader discourse on integrating cultural specificity into AI applications.

METHODS

Study Design and Data Collection

This study aimed to evaluate the accuracy of calorie information provided by three widely used chatbots-ChatGPT 4.0, Microsoft Copilot, and Gemini-for traditional Turkish foods [49-51]. A total of 71 foods were selected from the Türkomp National Food Composition Database [56], where they are specifically categorised as Traditional Turkish Foods (Table 1). The caloric content of each food item was obtained by individually querying each chatbot. The standardised question presented to the chatbots was: "What is the caloric content of [food name] per100 grams?". This question was presented to each chatbot in Turkish to ensure consistency across platforms. The decision to use Turkish as the input language was intentional, as it matches the native language context of the foods in question. Since these traditional dishes originate from Turkey, it was anticipated that chatbots might return more accurate and culturally relevant responses when prompted in Turkish. The calorie values returned by each chatbot were systematically recorded and compared against the reference values provided in the Türkomp database.

The prompt explicitly asked for the caloric content of each food per 100 grams only. However, some chatbots still provided answers in different formats such as per serving. In these cases, the responses were carefully reviewed and converted to a standard "per 100 grams" format to ensure accurate comparison with the Türkomp reference.

To keep the method consistent, all chatbot queries were performed on the same day and in the same order for each platform. This approach helped avoid issues such as internet interruptions or variation in chatbot responses over time, as chatbots may give different answers to the same question on different days.

This study did not involve human or animal subjects and used publicly accessible data. Therefore, ethical approval was not required.

Data Analysis

For each food item, the calorie accuracy percentage was calculated for each chatbot using the formula: Calorie Accuracy Percentage (%) = (1 - (|Chatbot Value - Reference Value| / Reference Value)) × 100

This formula expresses how closely the calorie values provided by each chatbot approached the reference value. A result of 100% indicates a perfect match, while lower percentages reflect increasing deviation. The subtraction from 1 ensures that the value represents accuracy rather than error. In cases where the calculated accuracy exceeded 100% or resulted in a negative value (indicating extreme overestimation or underestimation), the score was adjusted and recorded as 0%, based on the assumption that such large deviations represent inaccurate estimations. For example, the reference caloric value of Afyon pastırması is 209 kcal, but one chatbot returned a value of 600 kcal, resulting in an accuracy of –187%, which was thus recoded as 0%.

The calorie accuracy percentages were analysed using two key methods. Firstly, the mean accuracy percentage was calculated independently for each

Table 1. Traditional Turkish foods selected from the Türkomp National Food Composition Database for caloric accuracy analysis (n=71)

Food item (Turkish)	Culinary description (English)
Afyon pastırması	A regional variety of pastırma (cured beef) from Afyon, known for its rich flavour and thick çemen layer.
Antep baklavası, fıstıklı	Pistachio baklava from Gaziantep, recognised for its superior quality and rich taste.
Aşure	A sweet pudding made from grains, legumes, dried fruits, and nuts; also known as Noah's Pudding.
Ayran, yayık	A salty yogurt-based drink, churned using traditional methods for a foamy texture.
Badem ezmesi	Almond paste, often sweetened and shaped into small treats.
Baklava, cevizli	Baklava made with walnuts instead of pistachios.
Baklava, fıstıklı	A rich dessert made of filo layers, filled with pistachios and syrup.
Boza	A thick, slightly sour fermented drink made from bulgur or millet.
Bozdağ kestane şekeri	Candied chestnuts from the Bozdağ region.
Bulgur, pilavlık	Coarse bulgur used for making pilaf, a staple grain in Turkish cuisine.
Çökelek	A crumbly, low-fat fresh cheese made from fermented or boiled yogurt.
Çubuk turşusu, salatalık	Famous cucumber pickles from Çubuk district.
Dondurma, yanıksı	Slightly burnt-flavoured traditional Turkish ice cream with a smoky taste.
Döner, et, pişmiş	Cooked vertical rotisserie meat (usually lamb or beef), shaved for serving.
Edirne beyaz peyniri	White cheese from Edirne, softer and saltier than feta.
Erzincan tulum peyniri	A sharp, aged cheese from Erzincan, ripened in animal skin bags.
Eski kaşar, Kars	Aged hard yellow cheese from Kars, strong in flavour.
Ezine peyniri	A white brined cheese from Ezine, known for its creamy texture.
Güllaç, yaprak	Thin, starchy sheets soaked in milk and rose water, typically served during Ramadan.
Gümüşhane dut pestili	Mulberry fruit leather from Gümüşhane, famous for its natural sweetness.
Hardaliye	A fermented non-alcoholic drink made from grape juice, mustard seeds, and cherry leaves.
İskilip turşusu, salatalık	Traditional cucumber pickles from İskilip, with a unique regional flavour.
İzmit pişmaniyesi	Flossy candy made of sugar and flour, similar to cotton candy.
Karnavas dut pekmezi	Mulberry molasses from Karnavas, known for its deep flavour.
Kavurma et	Preserved cooked meat, usually fried and stored in its own fat.
Kayseri pastırması	Famous cured beef from Kayseri, thinner and with a distinct spicy coating.
Kayseri sucuğu	Famous spiced sausage from Kayseri.
Kazandibi	A caramelised milk pudding with a burnt bottom layer.
Kefir	A fermented milk drink rich in probiotics.
Keşkül	A creamy almond-based milk pudding, often garnished with nuts.
Kestane şekeri	Candied chestnuts, especially popular in Bursa.
Kuru incir	Dried figs, often eaten as a snack or used in desserts.
Leblebi	Roasted chickpeas, often served as a snack.
Lokum, kaymaklı	Turkish delight filled with clotted cream (kaymak).
Lokum, sade	Plain Turkish delight, typically dusted with powdered sugar.
Lokum, safranlı, antep fıstıklı	Turkish delight with saffron and pistachios.
Mantı, çiğ	Raw Turkish dumplings typically filled with ground meat and onions, to be boiled or steamed.

Table 1 continued. Traditional Turkish foods selected from the Türkomp National Food Composition Database for caloric accuracy analysis (n=71)

Food item (Turkish)	Culinary description (English)
Maraş dondurması	Stretchy, chewy ice cream from Maraş, made with salep and goat milk.
Maraş tarhanası	A fermented dried soup base made of yogurt, wheat, and spices from Maraş.
Mersin cezeryesi, cevizli	A sweet made from carrots and nuts (especially walnuts), typical in Mersin.
Muhallebi	A silky, sweet milk pudding, often served plain or with rose water.
Nar ekşisi	Pomegranate molasses; a tangy-sweet syrup used as a condiment or dressing.
Oltu cağ kebabı (döner, et, pişmiş)	Horizontally grilled meat skewers from Oltu, a regional döner variant.
Pastırma	Air-dried cured beef, seasoned with spices; a traditional delicacy.
Pastırma çemeni	A spicy garlic-based paste (usually with fenugreek) that coats pastırma (cured beef).
Pekmez, andız	Molasses made from the fruit of the wild terebinth tree (andız).
Pekmez, dut	Mulberry molasses, a natural syrup with a deep flavour.
Pekmez, harnup (keçiboynuzu)	Carob molasses; thick, sweet, and rich in minerals.
Pekmez, incir	Fig molasses, made by boiling fig juice into a thick syrup.
Pekmez, üzüm	Grape molasses; a natural sweetener made by boiling grape juice.
Pestil, dut	Thin fruit leather made from mulberries, dried into sheets.
Pestil, üzüm	Grape fruit leather, made by boiling and drying grape must.
Peynir, otlu	Herb cheese from Eastern Anatolia, made with local herbs.
Peynir, tulum	Aged cheese typically matured in goat skin (tulum).
Reçel, gül	Rose petal jam, known for its floral aroma and vibrant colour.
Salatalık turşusu	Pickled cucumber.
Salep, toz	Powder made from orchid tubers, used to prepare a warm, thick, sweet drink.
Simit	Circular sesame-crusted bread, commonly eaten as a street food.
Sucuk	Spicy, dry fermented sausage made with beef and garlic.
Sütlaç	Baked rice pudding made with milk, sugar, and rice.
Tahin	Sesame paste used in savoury and sweet dishes.
Tahin helvası	Dense sweet made from tahini and sugar, often eaten with bread.
Tarhana, kuru	Dried fermented soup powder made from yogurt, flour, and vegetables.
Tavşanlı leblebisi	Roasted chickpeas from Tavşanlı, known for their crunch and golden colour.
Tel kadayıf, çiğ	Raw shredded dough used for making sweet layered desserts.
Yaprak sarma, etli	Stuffed vine leaves with ground meat and rice.
Yaprak sarma, zeytinyağlı	Vine leaves stuffed with rice, pine nuts, and currants, cooked in olive oil.
Yaz helvası, cevizli	A summer-style helva with walnuts, lighter and more crumbly.
Yoğurt	Plain yogurt, a staple in Turkish cuisine and used in many dishes.
Yoğurt, süzme	Strained yogurt, thicker and creamier than regular yogurt.
Zile pekmezi (beyaz, sert, üzüm pekmezi)	White, firm grape molasses from Zile, eaten as a sweet.

chatbot to provide a comprehensive overview of their performance in estimating caloric values. Secondly, the accuracy percentages for all food items were classified into predefined intervals: ≥90%, 89.9%-80%, 79.9%-70%, 69.9%-60%, 59.9%-50%, and ≤49.9%. This categorisation allowed for a more detailed assessment of the distribution and variability in the chatbots' accuracy levels. For each chatbot, the frequency of

food items within each predefined accuracy range was determined and summarised to identify patterns in performance and precision. Additionally, foods for which all chatbots achieved a calorie accuracy of ≥90% were identified and reported.

Statistical Analysis

Statistical analysis was conducted to determine the mean accuracy percentages and identify differences among the chatbots. To evaluate differences in performance across the chatbots, a one-way Analysis of Variance (ANOVA) was conducted, comparing the mean accuracy percentages of the chatbots using the latest version (30.0.0) of IBM SPSS Statistics software. Statistical significance was set at $P< 0.05$.

RESULTS

The calorie accuracy of the three chatbots was assessed specifically for traditional Turkish foods, producing the following findings: ChatGPT 4.0 demonstrated a mean accuracy of 81.62%±20.6%, with a range of 12.50% to 99.78%. Similarly, Microsoft Copilot achieved a mean accuracy of 81.23%±20.7%, with estimates ranging from 0% to 100%. In comparison, Gemini exhibited a slightly lower mean accuracy of 70.99%±30.2%, ranging from 0% to 99.78%. A one-way ANOVA was conducted to examine whether there were significant differences in the mean accuracy percentages of the chatbots. The results revealed a statistically significant difference between the chatbots, $F (2, 210) = 4.39$, $P=0.0136$. Tukey's HSD test was used to perform pairwise comparisons between the three chatbots. The results showed that the differences in mean accuracy between ChatGPT 4.0 and Microsoft Copilot, ChatGPT 4.0 and Gemini, as well as Microsoft Copilot and Gemini, were all statistically significant ($P<0.05$). This indicates that each chatbot's mean accuracy was significantly different from the others.

The percentage of items with an accuracy equal to or greater than 90% varied significantly among the three chatbots. Gemini demonstrated a 90% or higher accuracy for 22 (30.9%) foods, while both Microsoft Copilot and ChatGPT 4.0 achieved this level of accuracy for 33 foods (46.5% each). Foods with an accuracy percentage equal to or less than 49.9% were more prevalent for Gemini, with 13 (18.3%) foods falling into this category, compared to 8 foods (11.3%) for Microsoft Copilot and 7 (9.8%) foods for ChatGPT 4.0. These findings illustrate significant variations in performance among the chatbots, with Gemini showing a higher proportion of low-accuracy estimates, while ChatGPT 4.0 and Microsoft Copilot demonstrated relatively better consistency in estimating caloric values with greater precision (Table 2, Fig. 1).

Eleven traditional Turkish foods achieved calorie accuracy percentages exceeding 90% across all three chatbots, underscoring their consistent performance in estimating caloric values for these items. These foods included kefir (fermented milk drink), Bozdağ chestnut candy, tahini halva (a sesame-based sweet), Eski Kaşar (aged cheese from Kars), Turkish delight flavoured with saffron and pistachios, summer halva with walnuts, Karnavas mulberry molasses, Zile molasses (white, hard, grape-based), mulberry molasses, pilaf-style bulgur wheat, and walnut baklava. Notably, five foods (buttermilk-style ayran, pomegranate molasses, uncooked Turkish dumplings, caramelised milk pudding, and güllaç sheets) were consistently identified among the chatbots as having low caloric estimation accuracy, with all three models reporting values below 49.9%.

DISCUSSION

The findings of this study reveal statistically significant differences in the performance of the evaluated

Table 2. Distribution of caloric value estimation accuracy across the three chatbots

	≥90%	89.9-80%	79.9-70%	69.9-60%	59.9-50%	≤49.9%
ChatGPT 4.0	n=33 (46.4%)	n=21 (29.5%)	n=3 (4.2%)	n=2 (2.8%)	n=5 (7.0%)	n=7 (9.8%)
Microsoft Copilot	n=33 (46.4%)	n=18 (25.3%)	n=3 (4.2%)	n=5 (7.0%)	n=4 (5.6%)	n=8 (11.2%)
Gemini	n=22 (30.9%)	n=20 (28.1%)	n=6 (8.4%)	n=4 (5.6%)	n=6 (8.4%)	n=13 (18.3%)

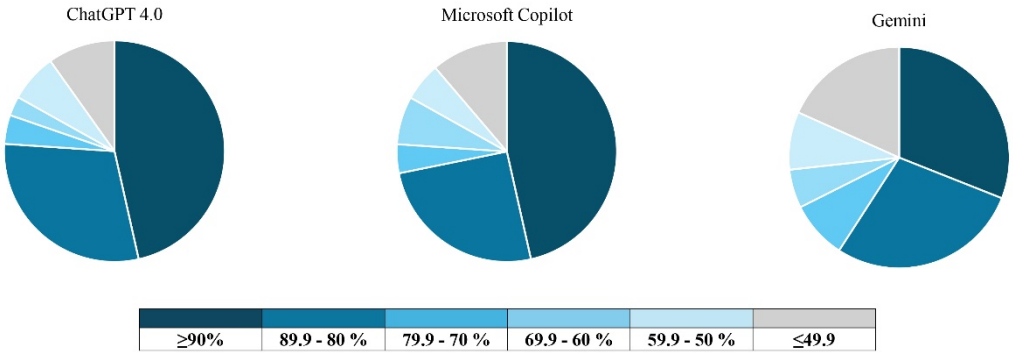

Fig. 1. Percentage distribution of food items by accuracy levels for ChatGPT 4.0, Microsoft Copilot, and Gemini.

chatbots in estimating caloric values. While ChatGPT 4.0 and Microsoft Copilot demonstrated higher consistency and precision, Gemini showed a higher percentage of low-accuracy estimates. These findings suggest that discrepancies in the underlying database quality or algorithmic capabilities may exist among the chatbots. Potential contributing factors may include differences in commercial agreements, the use of proprietary versus open-source datasets, and varying levels of access to specialised training data. Eleven traditional Turkish foods achieved calorie accuracy percentages exceeding 90% across all chatbots. The high agreement among chatbots for these foods suggests that their caloric values are well-represented in existing nutritional databases and demonstrates the potential reliability of chatbot-generated estimates for certain food categories. Another important factor may be the relatively simple or standardised nutrient composition of these foods, which likely facilitates more accurate caloric predictions. Compared to more complex dishes with variable recipes and portion sizes, foods with consistent ingredients and preparation methods may be easier for AI models to evaluate reliably. Consistently, the persistent misestimation of caloric values for five specific foods may present unique challenges for chatbot algorithms due to factors such as recipe variability, limited international exposure, and underrepresentation in global food composition databases. Additionally, the structural complexity or niche nature of some items (e.g., güllaç sheets used only in certain contexts) may contribute to inaccurate estimations.

Comparable accuracy levels of ChatGPT 4.0 and Microsoft Copilot may suggest potential similarities in their algorithms or access to diverse and comprehensive nutritional datasets. However, Gemini demonstrated slightly lower overall performance, with a mean accuracy of 70.99%. This discrepancy in accuracy among the AI tool may originate several factors, including differences in the underlying database quality (e.g., less comprehensive or outdated nutritional data), algorithmic capabilities (such as the use of less sophisticated learning architectures), and the frequency of model updates. Additionally, reliance on publicly available or regionally biased data sources could further limit effectiveness of chatbots in accurately identifying or categorising certain food items.

The high accuracy (≥90%) observed for several traditional Turkish foods across all chatbots might be explained by the presence of these items in widely accessible and detailed food composition databases. Foods such as kefir, tahini halva, and aged cheese from Kars (Eski Kaşar) are well-documented and popular both within and beyond Türkiye, making them more likely to be included in the datasets used to train these chatbots.

Previous studies have shown that AI systems, including chatbots, exhibit varying levels of accuracy depending on the complexity of the food item and the comprehensiveness of the underlying datasets [57-59]. In a study evaluating the performance of four chatbots (ChatGPT-3.5, ChatGPT-4.0, Bard AI, and Bing Chat) in categorising foods based on their oxalate content, the accuracy of the chatbots was found to decrease with a higher degree of dietary oxalate content categories [57]. Specifically, the study found that the accuracy of Bard AI, which had the highest overall accuracy, was 84% for low oxalate content foods, 79%

for moderate oxalate content foods, and 79% for high oxalate content foods [57]. These findings suggest that chatbots may struggle with more complex food items, such as those with high oxalate content [57]. Recent studies found that the chatbots' accuracy was also affected by the complexity of the food item, with some foods being more accurately identified by certain chatbots than others [60]. For example, the study found that ChatGPT-3.5 was more accurate in identifying foods with high phosphorus content, while Bard AI was more accurate in identifying foods with high potassium content [60]. Another research has demonstrated that chatbots, such as ChatGPT and Bard, can generate meal plans that are appropriate according to dietary reference intakes but often fail to achieve complete nutritional adequacy, particularly for restrictive dietary patterns like vegan diets [61]. In addition, a study evaluating the performance of chatbots in differentiating between medical emergency and non-emergency scenarios found that the chatbots tended to overclassify the scenarios as emergencies and underclassify them as non-emergencies [62]. This suggests that chatbots may also struggle with complex and nuanced scenarios, and that their accuracy may vary depending on the specific context. Overall, these findings suggest that AI systems, including chatbots, exhibit varying levels of accuracy depending on the complexity of the food item [6], and that their accuracy may vary depending on the specific food item and context [62, 63].

AI-based nutrition applications have reported similar challenges in accurately estimating the nutritional content of culturally specific or complex dishes [11, 47, 64-66]. These challenges include difficulties in identifying individual components within mixed dishes, variability of food appearance, type, and shape, and the inability to detect ingredients added in cooking or sauces and condiments [10, 65, 67, 68]. Additionally, AI models can struggle to accurately estimate portion sizes or volume, which can compound the inaccuracies associated with energy estimations [65, 69-71]. To address these challenges, researchers have proposed the use of multimodal foundation models, such as GPT-4V, which have demonstrated remarkable generalist intelligence and accuracy in dietary assessment tasks [72]. These models can use surrounding objects as scale references to estimate the portion sizes of food items and can be directed with specific language prompts to accurately identify regional dishes [72]. However, even with these advancements, AI-based nutrition applications still require improvement in their ability to generalise across a diverse range of food categories, dietary behaviours, and cultural contexts [6, 12, 14, 72, 73]. It is important to train models on diverse and region-specific datasets to improve their applicability in global contexts. The strong accuracy observed for widely recognised traditional Turkish foods, such as kefir and tahini halva, further supports existing literature indicating that AI tools tend to perform better with well-documented foods that are included in multiple international food composition databases. This study contributes to the growing body of evidence supporting the development of AI models that are both globally informed and locally adaptable, particularly for use in nutrition-related applications.

Strengths and Limitations

This study has several strengths that significantly enhance its contribution to the fields of nutrition science and artificial intelligence. First, it provides a focused evaluation of chatbot accuracy in estimating the caloric values of traditional Turkish foods, an underexplored area. By examining the performance of chatbots with region-specific foods, this research addresses an important gap and provides essential information for the development of culturally sensitive dietary assessment tools. Secondly, the study utilises a reliable and standardised food composition database, Türkomp, ensuring consistency and credibility in the reference values employed for comparison. Furthermore, the inclusion of three distinct chatbots-ChatGPT 4.0, Microsoft Copilot, and Gemini-allows for a comparative analysis of their performance, clarifying potential strengths and weaknesses in their algorithms and training datasets.

However, this study also has several limitations that should be acknowledged. First, the sample size of 71 foods, while adequate for providing initial understanding, is relatively small. This may limit the generalisability of the findings to other food items that differ in composition. Second, the use of a single reference database, Türkomp, as the gold standard for caloric values may have influenced the results. While Türkomp is a reliable and comprehensive resource for Turkish food composition data, the use of alternative

databases could produce different outcomes, especially for foods with varying preparation methods or regional variations. Third, parameters influencing LLM output variability such as temperature, maximum token length, and sampling strategies were not standardised across platforms, potentially contributing to inconsistencies in the generated responses. Lastly, this evaluation did not include a qualitative or quantitative analysis of hallucination phenomena, nor did it assess whether the chatbot responses were based on verifiable sources or referenced nutritional data, which limits interpretability in terms of factual grounding.

The suboptimal performance observed in this study raises an important question: are these limitations inherent to the language models themselves, or do they reflect a fundamental mismatch between the task and the capabilities for which LLMs are designed? LLMs are optimised for language generation and reasoning based on patterns in large-scale text corpora, rather than precise quantitative estimations grounded in scientific data. LLMs might lack true understanding of the factual world, operating instead on statistical correlations rather than structured knowledge [74, 75]. This becomes particularly problematic in domains like nutrition, where accurate caloric estimation requires not just textual fluency but access to verifiable, domain-specific databases [76]. Moreover, prompt sensitivity and model behaviour can vary drastically depending on input phrasing and hidden inference parameters, which complicates reproducibility and reliability [77, 78]. Our results thus reflect both a limitation of current LLM architectures and a task misalignment and highlight the need for hybrid approaches that combine LLMs with structured, domain-specific knowledge bases.

Building on these findings, future research should explore a larger and more diverse dataset encompassing foods from multiple cuisines and preparation styles. This would enable a more comprehensive assessment of chatbot performance across global dietary contexts. Additionally, expanding the scope of evaluation to include macronutrient and micronutrient estimation would provide a deeper understanding of chatbot utility in nutritional science. Furthermore, research on user experience, trust, and the perceived reliability of chatbot-generated nutritional information is essential to understanding the broader implications of using these tools in real-world settings.

CONCLUSION

By evaluating the performance of ChatGPT 4.0, Microsoft Copilot, and Gemini in estimating the caloric values of traditional Turkish foods, the findings contribute to the growing body of research on AI applications in nutrition. The results demonstrate both the potential and challenges of current chatbot technologies and emphasise the need for improved algorithms and culturally diverse food databases. This research serves as a stepping stone for further studies, contributing to the development of more accurate and reliable AI-driven tools in dietary assessment and public health nutrition. However, it is important to acknowledge the ongoing role of human expertise in interpreting and validating these tools, as human judgment is essential for addressing the complex, context-specific factors that AI systems may not fully capture. Looking ahead, this work offers a wider perspective for future research by providing a foundation for the responsible integration of chatbot technologies into clinical nutrition practice and public health initiatives. Interdisciplinary collaboration between AI developers, nutritional scientists, and healthcare practitioners will be crucial for optimising chatbot design and functionality to meet the diverse demands of real world applications.

Ethics Approval and Consent to Participate

This study did not involve human or animal subjects and used publicly accessible data. Therefore, Ethics Committee approval was not required. This study does not require informed consent.

Data Availability

The main data generated or analysed during this study are included in this published article. Additional data that support the findings of this study are available from the corresponding author upon reasonable request.

Authors' Contribution

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

Conflict of Interest

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

Financing

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

Acknowledgments

This study has not been published as part of any thesis or project.

Generative Artificial Intelligence Statement

The authors declare that no artificial intelligence-based tools or applications were used in the preparation of this manuscript. ChatGPT 4.0, Microsoft Copilot, and Gemini were evaluated solely as research subjects, and all content was produced by the authors in accordance with scientific research methods and academic ethical principles.

Editor's Note

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

REFERENCES

1. Laranjo L, Dunn AG, Tong HL, et al. Conversational agents in healthcare: a systematic review. J Am Med Inform Assoc. 2018;25(9):1248-1258. doi: 10.1093/jamia/ocy072.
2. Xu L, Sanders L, Li K, Chow JCL. Chatbot for Health Care and Oncology Applications Using Artificial Intelligence and Machine Learning: Systematic Review. JMIR Cancer. 2021;7(4):e27850. doi: 10.2196/27850.
3. Kulkarni R, Jaiswal CR. A survey on AI Chatbots. Int J Res Appl Sci Eng Technol. 2023;11(9):1738-1744. doi: 10.22214/ijraset.2023.55911.
4. Tsolakidis D, Gymnopoulos LP, Dimitropoulos K. Artificial Intelligence and Machine Learning Technologies for Personalized Nutrition: A Review. Informatics 2024;11(3):62. doi: 10.3390/informatics11030062
5. Abeltino A, Riente A, Bianchetti G, et al. Digital applications for diet monitoring, planning, and precision nutrition for citizens and professionals: a state of the art. Nutr Rev. 2025;83(2):e574-e601. doi: 10.1093/nutrit/nuae035.
6. Zhang J, Oh YJ, Lange P, Yu Z, Fukuoka Y. Artificial Intelligence Chatbot Behavior Change Model for Designing Artificial Intelligence Chatbots to Promote Physical Activity and a Healthy Diet: Viewpoint. J Med Internet Res. 2020;22(9):e22845. doi: 10.2196/22845.
7. Johnson KB, Wei WQ, Weeraratne D, et al. Precision Medicine, AI, and the Future of Personalized Health Care. Clin Transl Sci. 2021;14(1):86-93. doi: 10.1111/cts.12884.
8. Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database (Oxford). 2020;2020:baaa010. doi: 10.1093/database/baaa010.
9. Budhwar P, Chowdhury S, Wood G, et al. Human resource management in the age of generative artificial intelligence: Perspectives and research directions on ChatGPT. Hum Resour Manag J. 2023;33(3):606-659. doi: 10.1111/1748-8583.12524.
10. Detopoulou P, Voulgaridou G, Moschos P, et al. Artificial intelligence, nutrition, and ethical issues: A mini-review. Clin Nutr Open Sci. 2023;50:46-56. doi: 10.1016/j.nutos.2023.07.001.
11. Côté M, Lamarche B. Artificial intelligence in nutrition research: perspectives on current and future applications. Appl Physiol Nutr Metab. 2022;47(1):1-8. doi: 10.1139/apnm-2021-0448.
12. Ponzo V, Goitre I, Favaro E, et al. Is ChatGPT an Effective Tool for Providing Dietary Advice? Nutrients. 2024;16(4):469. doi: 10.3390/nu16040469.
13. Hoang YN, Chen YL, Ho DKN, et al. Consistency and Accuracy of Artificial Intelligence for Providing Nutritional Information. JAMA Netw Open. 2023;6(12):e2350367. doi: 10.1001/jamanetworkopen.2023.50367.
14. Liao LL, Chang LC, Lai IJ. Assessing the Quality of ChatGPT's Dietary Advice for College Students from Dietitians' Perspectives. Nutrients. 2024;16(12):1939. doi: 10.3390/nu16121939.
15. Wang L, Wan Z, Ni C, et al. Applications and Concerns of ChatGPT and Other Conversational Large Language Models in Health Care: Systematic Review. J Med Internet Res. 2024;26:e22769. doi: 10.2196/22769.
16. Chew HSJ. The Use of Artificial Intelligence-Based Conversational Agents (Chatbots) for Weight Loss: Scoping Review and Practical Recommendations. JMIR Med Inform. 2022;10(4):e32578. doi: 10.2196/32578.
17. Thirunavukarasu AJ, Ting DSJ, Elangovan K, Gutierrez L, Tan TF, Ting DSW. Large language models in medicine. Nat Med. 2023;29(8):1930-1940. doi: 10.1038/s41591-023-02448-8.
18. Fadhil A, Gabrielli S. Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. PervasiveHealth '17: Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare. 2017:261-265. doi: 10.1145/3154862.3154914.
19. Bonan M, Ross H, et al. Recent Advances in Natural Language Processing via Large Pre-trained Language Models: A Survey. ACM Computing Surveys. 2023;56(2):1-40. doi: 10.1145/3605943.
20. Hagos DH, Battle R, Rawat DB. Recent Advances in Generative AI and Large Language Models: Current Status, Challenges, and Perspectives. IEEE Trans Artif Intell. 2024;5(2):5873-5893. doi: 10.1109/TAI.2024.3444742.
21. Hadi MU, Qureshi R, Shah A, Irfan M, Zafar A, Shaikh MB, et

al. Large language models: a comprehensive survey of its applications, challenges, limitations, and future prospects. Authorea preprints. 2023;1(3):1-26. https://www.techrxiv.org/users/618307/articles/682263-large-language-models-a-comprehensive-survey-of-its-applications-challenges-limitations-and-future-prospects.
22. Al-Surmi A, Bashiri M, Koliousis I. AI based decision making: combining strategies to improve operational performance. Int J Prod Res. 2022;60(14):4464-4486. doi: 10.1080/00207543.2021.1966540.
23. Khlaif ZN, Mousa A, Hattab MK, et al. The Potential and Concerns of Using AI in Scientific Research: ChatGPT Performance Evaluation. JMIR Med Educ. 2023;9:e47049. doi: 10.2196/47049.
24. Umoga UJ, Sodiya EO, Ugwuanyi ED, et al. Exploring the potential of AI-driven optimization in enhancing network performance and efficiency. Magna Scientia Advanced Research and Reviews. 2024;10(1):368-78. doi: 10.30574/msarr.2024.10.1.0028
25. Wang C, Liu X, Yue Y, et al. Survey on factuality in large language models: Knowledge, retrieval and domain-specificity. arXiv preprint arXiv:231007521. 2023. doi: 10.48550/arXiv.2310.07521.
26. Huang Y, Tang K, Chen M, Wang B. A comprehensive survey on evaluating large language model applications in the medical industry. arXiv preprint arXiv:240415777. 2024. doi: 10.48550/arXiv.2404.15777
27. Annepaka Y, Pakray P. Large language models: a survey of their development, capabilities, and applications. Knowl Inf Syst. 2025;67(3):2967-3022. doi: 10.1007/s10115-024-02310-4
28. Giannakopoulos K, Kavadella A, Aaqel Salim A, Stamatopoulos V, Kaklamanos EG. Evaluation of the performance of generative AI large language models ChatGPT, Google Bard, and Microsoft Bing Chat in supporting evidence-based dentistry: comparative mixed methods study. J Med Internet Res. 2023;25:e51580. doi: 10.2196/51580
29. Gichoya JW, Banerjee I, Bhimireddy AR, et al. AI recognition of patient race in medical imaging: a modelling study. Lancet Digit Health. 2022;4(6):e406-e414. doi: 10.1016/S2589-7500(22)00063-2
30. Collenette J, Atkinson K, Bench-Capon T. Explainable AI tools for legal reasoning about cases: A study on the European Court of Human Rights. Artificial Intelligence. 2023;317:103861. doi: 10.1016/j.artint.2023.103861.
31. Zhang C, Meng Y. Bayesian deep learning: An enhanced AI framework for legal reasoning alignment. Comput Law Security Rev. 2024;55:106073. doi: 10.1016/j.clsr.2024.106073.
32. Spaić B, Jovanović M. Artificial Reason and Artificial Intelligence: The Legal Reasoning Capabilities of GPT-4. Annals of the Faculty of Law in Belgrade. 2024;72(3). doi: 10.51204/Anali_PFBU_24302A.
33. Shanmuganathan M. Behavioural finance in an era of artificial intelligence: Longitudinal case study of robo-advisors in investment decisions. J Behav Exp Finance. 2020;27:100297. doi: 10.1016/j.jbef.2020.100297.
34. Sultan MS, Sultan MS. Enhancing Corporate Financial Performance Through AI: A Novel AI Model for Forecasting Organizational Risk Management, CRM, and Operational Efficiency. Inter J Sci Res (IJSR). 2024;13(8):1758-1765. doi: 10.21275/SR24828091042.
35. Carels RA, Harper J, Konrad K. Qualitative perceptions and caloric estimations of healthy and unhealthy foods by behvioral weight loss participants. Appetite. 2006;46(2):199-206. doi: 10.1016/j.appet.2005.12.002.
36. Wang Z, Wang L, Hou Y, et al. Precision calories: A promising strategy for personalized health interventions in the precision nutrition framework. Trends Food Sci Technol. 2024;153:104727. doi: 10.1016/j.tifs.2024.104727.
37. Almasi N, Rakıcıoğlu N. The Estimation of Standard Portion Sizes Using Food Photographic Booklet Among University Students in Turkey. Clin Exp Health Sci 2021;11(4):775-781. doi: 10.33808/clinexphealthsci.886274.
38. Çakmak M, Sarıışık M. An Investigation on the Basic Contents of the Main Dishes of the Turkish Cuisine. Rev Anais Bras de Est Tur (ABET). 2019;9(1, 2 e3):1-13. doi: 10.34019/2238-2925.2019.v9.27149.
39. Biringen Loker G, Ozkoc SO, Amoutzopoulos B, et al. Establishing a food composition database for Turkey based on European standards. Nutr Bull. 2011;36(2):254-258. doi: 10.1111/j.1467-3010.2011.01898.x.
40. Biringen Löker G, Amoutzopoulos B, Özge Özkoç S, Özer H, Şatir G, Bakan A. A pilot study on food composition of five Turkish traditional foods. Br Food J. 2013;115(3):394-408. doi: 10.1108/00070701311314219.
41. Kirk D, Catal C, Tekinerdogan B. Precision nutrition: A systematic literature review. Comput Biol Med. 2021;133:104365. doi: 10.1016/j.compbiomed.2021.104365.
42. Gami SJ, Dhamodharan B, Dutta PK, Gupta V, Whig P. Data Science for Personalized Nutrition Harnessing Big Data for Tailored Dietary Recommendations. In: Gami SJ, Dutta PK, Whing, eds., Nutrition Controversies and Advances in Autoimmune Disease. IGI Global Publishing. 2024: pp. 606-630.
43. Prabhakaran V, Qadri R, Hutchinson B. Cultural incongruencies in artificial intelligence. arXiv preprint arXiv:2211.13069. 2022. doi: 10.48550/arXiv.2211.13069.
44. Zhou L, Karidi T, Liu W, et al. Does mapo tofu contain coffee? probing llms for food-related cultural knowledge. arXiv preprint arXiv:24040.6833. 2024. doi: 10.48550/arXiv.2404.06833.
45. Winata GI, Hudi F, Irawan PA, et al. Worldcuisines: A massive-scale benchmark for multilingual and multicultural visual question answering on global cuisines. arXiv preprint arXiv:24101.2705. 2024. doi: 10.48550/arXiv.2410.12705.
46. Amoutzopoulos B, Page P, Roberts C, et al. Portion size estimation in dietary assessment: a systematic review of existing tools, their strengths and limitations. Nutr Rev. 2020;78(11):885-900. doi: 10.1093/nutrit/nuz107.
47. Abdusalomov A, Mukhiddinov M, Djuraev O, Khamdamov U, Abdullaev U, editors. AI-Based Estimation from Images of Food Portion Size and Calories for Healthcare Systems. International Conference on Intelligent Human Computer Interaction; 2023: Springer. doi: 10.1007/978-3-031-53830-8_2.
48. Konstantakopoulos FS, Georga EI, Fotiadis DI. A review of image-based food recognition and volume estimation artificial intelligence systems. IEEE Rev Biomed Eng. 2023;17:136-52. doi: 10.1109/RBME.2023.3283149.
49. ChatGPT 4.0. Open AI ChatGPT. Availabe online: https://chatgpt.com/ (accessed on 20.11.2024).
50. Copilot. Microsoft Copilot. Availabe online:

https://copilot.microsoft.com/ (accessed on 23.11.2024).
51. Gemini. Google Gemini App. Availabe online: https://gemini.google.com/app (accessed on 10.11.2024).
52. Roumeliotis KI, Tselikas ND. Chatgpt and open-ai models: A preliminary review. Future Internet. 2023;15(6):192. doi: 10.3390/fi15060192.
53. Jaffe S, Parikh N, Butler JL, Farach A, Cambon A, Hecht B, et al., editors. Generative AI in Real-World Workplaces: The Second Microsoft Report on AI and Productivity Research. 2024.
54. Šarac M, editor Microsoft Copilot as a Transformative Tool in Business: Opportunities and Challenges 2025: Singidunum University. doi: 10.15308/Sinteza-2025-347-353.
55. Akhtar ZB. From bard to Gemini: An investigative exploration journey through Google's evolution in conversational AI and generative AI. Comput Artif Intell. 2024;2(1):1378. doi: 10.59400/cai.v2i1.1378.
56. Database, N.F.C. Turkish. Availabe online: https://turkomp.tarimorman.gov.tr/main (accessed on 10.11.2024).
57. Aiumtrakul N, Thongprayoon C, Arayangkool C, et al. Personalized Medicine in Urolithiasis: AI Chatbot-Assisted Dietary Management of Oxalate for Kidney Stone Prevention. J Pers Med. 2024;14(1):107. doi: 10.3390/jpm14010107.
58. Bayram HM, Ozturkcan SA. AI showdown: info accuracy on protein quality content in foods from ChatGPT 3.5, ChatGPT 4, bard AI and bing chat. Br Food J. 2024;126(9):3335-3346. doi: 10.1108/bfj-02-2024-0158.
59. Naja F, Taktouk M, Matbouli D, et al. Artificial intelligence chatbots for the nutrition management of diabetes and the metabolic syndrome. Eur J Clin Nutr. 2024;78(10):887-896. doi: 10.1038/s41430-024-01476-y.
60. Qarajeh A, Tangpanithandee S, Thongprayoon C, et al. AI-Powered Renal Diet Support: Performance of ChatGPT, Bard AI, and Bing Chat. Clin Pract. 2023;13(5):1160-1172. doi: 10.3390/clinpract13050104.
61. Hieronimus B, Hammann S, Podszun MC. Can the AI tools ChatGPT and Bard generate energy, macro- and micro-nutrient sufficient meal plans for different dietary patterns? Nutr Res. 2024;128:105-114. doi: 10.1016/j.nutres.2024.07.002.
62. Salazar GZ, Zúñiga D, Vindel CL, et al. Efficacy of AI Chats to Determine an Emergency: A Comparison Between OpenAI's ChatGPT, Google Bard, and Microsoft Bing AI Chat. Cureus. 2023;15(9):e45473. doi: 10.7759/cureus.45473.
63. Safitri S, Mantoro T, Bhakti MAC, Wandy W. Cooking and Food Information Chatbot System using GPT-3. In Proceedings of 2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED); 2023:1-6. doi: 10.1109/ICCED60214.2023.10425553.
64. Cathcart A, Wiggins M, Kelly W, et al. Recognition of Culturally Diverse Cuisine in Popular Nutrition Apps. Curr Dev Nutr. 2022;6:872. doi: 10.1093/cdn/nzac066.002.
65. Li X, Yin A, Choi HY, Chan V, Allman-Farinelli M, Chen J. Evaluating the Quality and Comparative Validity of Manual Food Logging and Artificial Intelligence-Enabled Food Image Recognition in Apps for Nutrition Care. Nutrients. 2024;16(15):2573. doi: 10.3390/nu16152573.
66. Thames Q, Karpur A, Norris W, et al. Nutrition5k: Towards Automatic Nutritional Understanding of Generic Food. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021, pp. 8899-8907. doi: 10.1109/CVPR46437.2021.00879.
67. Yang Z, Khatibi E, Nagesh N. ChatDiet: Empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework. Smart Health. 2024; 32(1):100465. doi: 10.1016/j.smhl.2024.100465.
68. Fadhil A. Can a chatbot determine my diet?: Addressing challenges of chatbot application for meal recommendation. ArXiv Preprint ArXiv:1802.09100. 2018. doi: 10.48550/arXiv.1802.09100.
69. Shonkoff E, Cara KC, Pei X, et al. The State of the Science on Artificial Intelligence-Based Dietary Assessment Methods That Use Digital Images: A Scoping Review. Curr Dev Nutr. 2022;6(Suppl 1):534. doi: 10.1093/cdn/nzac077.037.
70. Limketkai BN, Mauldin K, Manitius N, Jalilian L, Salonen BR. The Age of Artificial Intelligence: Use of Digital Technology in Clinical Nutrition. Curr Surg Rep. 2021;9(7):20. doi: 10.1007/s40137-021-00297-3.
71. Sharma SK, Gaur S. Optimizing Nutritional Outcomes: The Role of AI in Personalized Diet Planning. Int J Res Publication Sem. 2024;15(2)::107-116. doi: 10.36676/jrps.v15.i2.15.
72. Lo FP, Qiu J, Wang Z, et al. Dietary Assessment with Multimodal ChatGPT: A Systematic Analysis. IEEE J Biomed Health Inform. 2024;28(12):7577-7587. doi: 10.1109/JBHI.2024.3417280.
73. Haman M, Školník M, Lošťák M. AI dietician: Unveiling the accuracy of ChatGPT's nutritional estimations. Nutrition. 2024;119:112325. doi: 10.1016/j.nut.2023.112325.
74. Hu X, Chen J, Li X, et al. Do large language models know about facts? arXiv preprint arXiv:2310.05177. 2023. doi: 10.48550/arXiv.2310.05177.
75. Venkatasubramanian V. Do large language models "understand" their knowledge? AIChE J. 2025;71(3):e18661. doi: 10.1002/aic.18661.
76. Azimi I, Qi M, Wang L, Rahmani AM, Li Y. Evaluation of LLMs accuracy and consistency in the registered dietitian exam through prompt engineering and knowledge retrieval. Sci Rep. 2025;15(1):1506. doi: 10.1038/s41598-024-85003-w.
77. Sclar M, Choi Y, Tsvetkov Y, Suhr A. Quantifying Language Models' Sensitivity to Spurious Features in Prompt Design or: How I learned to start worrying about prompt formatting. arXiv preprint arXiv:2310.11324. 2023. doi: 10.48550/arXiv.2310.11324.
78. Guan B, Roosta T, Passban P, Rezagholizadeh M. The Order Effect: Investigating Prompt Sensitivity to Input Order in LLMs. arXiv preprint arXiv:2502.04134. 2025. doi: 10.48550/arXiv.2502.04134.

The European Research Journal 2025;11(5):934-943 Original Article

DOI: https://doi.org/10.18621/eurj.1745159 *Pediatric Neurology*

The role of lipid profile and inflammotry markers in diffrentiating between focal and generalized types of seizures in pediatric epilepsy patients

Serap Bilge[1], Hüseyin Bilgin[2]

[1]Department of Pediatric Neurology, Ministry of Health, Diyarbakır Children's Hospital, Diyarbakır, Türkiye; [2]Department of Pediatric Metabolism, Ministry of Health, Diyarbakır Children's Hospital, Diyarbakır, Türkiye

ABSTRACT

Objectives: Seizures result from excessive and abnormal neuronal excitation, accompanied by pathophysiological processes such as inflammation, oxidative stress, and apoptotic cell death. Previous studies have demonstrated that patients with epilepsy often exhibit dysregulated inflammatory, vascular, and metabolic pathways, with certain seizure types associated with high pro-inflammatory cytokines. We aimed to investigate whether lipid profile parameters and inflammatory markers can be used to differentiate between focal and generalized seizures in pediatric epilepsy patients.

Methods: A total of 100 pediatric epilepsy patients and 100 age- and sex-matched healthy controls were recruited at Diyarbakır Children's Hospital between December 2021 and March 2023. Ethical approval was obtained from the institutional review board, and informed consent was secured from parents/guardians. Lipid levels (triglycerides, total cholesterol, low-density lipoprotein [LDL] cholesterol, high-density lipoprotein [HDL]) cholesterol) and inflammatory markers (ferritin, glucose, white blood cell count [WBC], C-reactive protein [CRP]) were analyzed using standard laboratory methods.

Results: The mean age was 9.2±4.4 years in the epilepsy group and 8.9±4.9 years in controls. Triglycerides, LDL, and total cholesterol levels were significantly higher in epilepsy patients compared to controls ($P<0.05$). Similarly, ferritin, glucose, WBC count, and CRP were elevated in the epilepsy group. Among seizure subtypes, patients with generalized seizures demonstrated significantly higher triglyceride, LDL, and total cholesterol levels compared with those with focal seizures ($P<0.05$). The WBC count was the only inflammatory marker that was elevated considerably in the generalized seizure group.

Conc**lusions:** Pediatric patients with epilepsy, particularly those with generalized seizures, exhibit higher lipid and inflammatory marker levels compared to controls and patients with focal seizures. These findings suggest potential diagnostic and prognostic roles for metabolic and inflammatory markers in seizure classification.

Keywords: Epilepsy, seizures, lipid profile, inflammatory markers, pediatrics

Seizures are characterized by abnormal and excessive neuronal activity and can be brought on by a complex web of pathophysiological processes, including increased production of reactive oxygen species, inflammation, and apoptotic cell death. People with persistent epilepsy have dysregulated vascu-

Received: July 17, 2025 **Accepted:** August 24, 2025 **Available Online:** August 26, 2025 **Published:** September 4, 2025

How to cite this article: Bilge S, Bilgin H. The role of lipid profile and inflammotry markers in diffrentiating between focal and generalized types of seizures in pediatric epilepsy patients. Eur Res J. 2025;11(5):934-943. doi: 10.18621/eurj.1733954

Corresponding author: Serap Bilge, MD., Phone: +90 412 224 57 51, E-mail: sarabsmustafa@hotmail.com

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

lar, metabolic, and inflammatory systems, according to earlier research. Interestingly, pro-inflammatory cytokine levels have been explicitly associated with specific seizure types, suggesting that distinct inflammatory pathways may contribute to seizure semiology [1-2].

In line with this, recent studies have demonstrated that lipid profiles and systemic inflammatory markers are interrelated, with dyslipidemia often accompanied by heightened inflammatory responses, thereby reinforcing the biological plausibility of our findings in pediatric epilepsy patients [1-4]. The objective of our study is to evaluate the relationship among lipid profiles, inflammatory markers, and seizures.

METHODS

A retrospective analysis of medical data from December 2021 to March 2023 at Diyarbakır Children's Hospital, revealing the following ICD-10 codes: G40.0 epilepsy, G40.1 epilepsy, G41.2 complex partial epilepsy, G41.8 other epilepsy, G41.9 unspecified epilepsy, and G40.7 optimal and generalized tonic-clonic seizures, was conducted. Data were collected regarding age, gender, weight, body mass index (BMI), seizure type (as determined by semiology and EEG findings), cause (if known), magnetic resonance imaging (MRI), electroencephalogram (EEG), hemogram, biochemical test, and lipid profile. Significant exclusion criteria included a history of autoimmune, liver, kidney, or inflammatory diseases, allergic reactions, immune deficiencies, diabetes, psychiatric disorders, cancer, smoking, systemic, metabolic, or central nervous system (CNS) infection within two weeks of sample collection, and elevated lipid levels 4-6 months after the first seizure. Furthermore, patients having a BMI of more than 25 or less than 18 were excluded from the study.

Seizure type and etiology were assessed using International League Against Epilepsy (ILAE) criteria. Written informed consent was obtained from all participants and their legal guardians. The study was conducted in accordance with the ethical principles outlined in the Declaration of Helsinki and was approved by the Ethics Committee of Health Sciences University Gazi Yaşargil Training and Research Hospital on August 4, 2023, with approval number 492.

Epilepsy Group

The seizure type for the epilepsy group (n=100) was confirmed through interviews with the patients

Table 1. The clinical and demographic profile of the patients and control group

	Epilepsy group (n=100)	Control group (n=100)	P value
Age (years)	9.2±4.4 9 (6-12.5)	8.9±4.9 9 (4-13)	0.605
Gender, n (%)			0.572
Male	53 (53%)	49 (%49%)	
Female	47 (47%)	51 (51%)	
Weight (kg)	39.3±16.4 40.5 (25.5-52.5)	36.2±14.4 37.5 (21.5-45.5)	0.157
Height (cm)	133.5±29.2 139 (109-158)	126±27.3 128 (101.5-146.5)	0.063
BMI (kg/m^2)	21.1±2.6 20.3 (19.2-22.7)	21.9±1.9 22.2 (20.2-23.8)	0.061
Epilepsy type, n (%)			
Focalized	46 (%46)		
Generalized	54 (%54)		

Data are shown as mean±standard deviation or median (IQR) where appropriate. BMI=Body mass index

and their relatives (cell phone videos), as well as EEG analysis. Patients were only included in the study when EEG findings were consistent with epilepsy. Specifically, EEG demonstrated focal epileptic activity in patients with focal seizures and generalized epileptic activity in those with generalized seizures. MRI was conducted in all patients, and only those with normal MRI findings were included to avoid the influence of other factors on the results. Furthermore, all patients had no identified etiology (genetic tests were not performed because the seizures were controlled with one or two medications).

Control Group

The control group consisted of 100 healthy volunteers with the same gender, BMI, and age.

Laboratory Analyses

Blood tests for glucose, CRP, and WBC were performed soon after admission. Total cholesterol (TC), low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides (TGL), and ferritin levels were all measured at least seven days after the first seizure, following an 8-12 hour overnight fast [5]. These values were collected from the patients' medical records. According to our treatment protocol, all patients who have their first seizure are thoroughly checked for any underlying metabolic issues.

Statistical Analysis

Categorical variables were expressed as numbers and percentages, whereas continuous variables were summarized as mean and standard deviation, and median and IQR where appropriate. To compare the distribution of gender between the groups and epilepsy types, the Pearson Chi-Square Test or Fisher's Exact Test was used, depending on whether the expected value problem arises or not. The normality of distribution for continuous variables was confirmed with the Shapiro-Wilk test. For comparison of continuous variables (age, BMI, lipid levels, and WBC, CRP, ferritin, and glucose) between the groups and the epilepsy types, the Student's t-test or Mann-Whitney U test was

Table 2. Lipid profile/ inflammatory markers differences between pediatric epilepsy patients and healthy controls

	Epilepsy group (n=100)	Control group (n=100)	P value
TC (mg/dL)	183.1±54.7 169.5 (135.5-216.5)	128.1±14.8 125 (117-141)	**<0.001**
TGL (mg/dL)	205.2±77.3 169 (143-267)	159.3±21.1 159 (145-176)	**<0.001**
LDL (mg/dL)	130.9±22.9 127 (111-145)	94.3±6.5 95 (88.5-98)	**<0.001**
HDL (mg/dL)	43.9±3.4 43 (42-46)	48.5±12.6 45 (45-46)	**<0.001**
WBC ($\times 10^3/\mu L$)	12.4±1.1 12 (12-13)	6.8±2.3 6.7 (5.7-6.8)	**<0.001**
CRP (mg/dL)	8.5±3.1 8 (6-11)	3.9±1 3 (3-5)	**<0.001**
Ferritin (ng/mL)	163.1±46 156 (143-176)	48.5±9.9 53 (43-56)	**<0.001**
Glucose (mg/dL)	82.3±9.7 84 (76-89)	83.4±4.7 85 (79-87)	0.284

Data are shown as mean±standard deviation or median (IQR) where appropriate. TC=Total cholesterol, TGL=Triglycerides, LDL=Low-density lipoprotein, HDL=High-density lipoprotein cholesterol, WBC=White blood cell, CRP=C reactive protein

used depending on whether the statistical hypotheses were fulfilled or not. All analyses were performed using the IBM SPSS Statistics Version 20.0 statistical software package. The statistical level of significance for all tests was considered to be 0.05.

RESULTS

A total of 100 pediatric epilepsy patients (54 with generalized seizures and 46 with focal seizures) and 100 healthy controls were included. The mean age, gender distribution, weight, height, and BMI did not differ significantly between the epilepsy and control groups (Table 1). Within the epilepsy cohort, demographic and anthropometric variables were also similar between the focal and generalized subgroups (Table 3).

Epilepsy patients exhibited significantly higher serum TC, TGL, and LDL levels compared with healthy controls, whereas HDL levels were significantly lower (all P<0.001) (Table 2; Figs. 1A-D). Specifically, total cholesterol (Fig. 1A), triglycerides (Fig. 1B), and LDL (Fig. 1D) were elevated, while HDL values (Fig. 1C) were reduced in the epilepsy group.

Markers of systemic inflammation were elevated in the epilepsy group compared with controls. Specifically, WBC count (Fig. 2A), CRP (Fig. 2B), and ferritin (Fig. 2C) levels were significantly higher in epilepsy patients (P<0.001 for all), while glucose levels (Fig. 2D) did not differ significantly between groups (P=0.284) (Table 2).

Children with generalized seizures had significantly higher TC (Fig. 3A), TGL (Fig. 3B), and LDL (Fig. 3D) levels compared with those with focal seizures (P<0.001), whereas HDL values (Fig. 3C) were similar across groups (Table 4). Among inflammatory markers, only WBC count (Fig. 4A) was significantly higher in the generalized group (P=0.031), while CRP (Fig. 4C), ferritin (Fig. 4D), and glucose levels (Fig. 4B) did not differ significantly between patients with focal and generalized seizures (Table 4).

DISCUSSION

In this research, it was noticed that the patients with seizures had increased levels of inflammation and glycolipids, particularly in the case of generalized seizures. The development and progression of seizures and epilepsy are more associated with such underlying oxidative and inflammatory processes in the brain [1]. To the best of our knowledge, this is the first research highlighting the importance of ongoing monitoring of

Table 3. Comparison of clinical variables between pediatric epilepsy patients with focal and generalized seizures

	Seizure type		P value
	Focal (n=46)	Generalized (n=54)	
Age (years)	8.7±4.5 9.5 (4-12)	9.7±4.3 9 (6-14)	0.227
Gender, n (%)			0.579
Male	23 (%50)	30 (%56)	
Female	23 (%50)	24 (%44)	
Weight (kg)	37.2±15.6 41 (19-51)	41.1±17 40.5 (26-54)	0.232
Height (cm)	130±30.2 139.5 (95-152)	136.5±28.2 137 (110-165)	0.272
BMI (kg/m^2)	21±2.3 20.5 (19.2-22.7)	21.1±2.9 20.3 (19.1-22.6)	0.881

Data are shown as mean±standard deviation or median (IQR) where appropriate. BMI=Body mass index

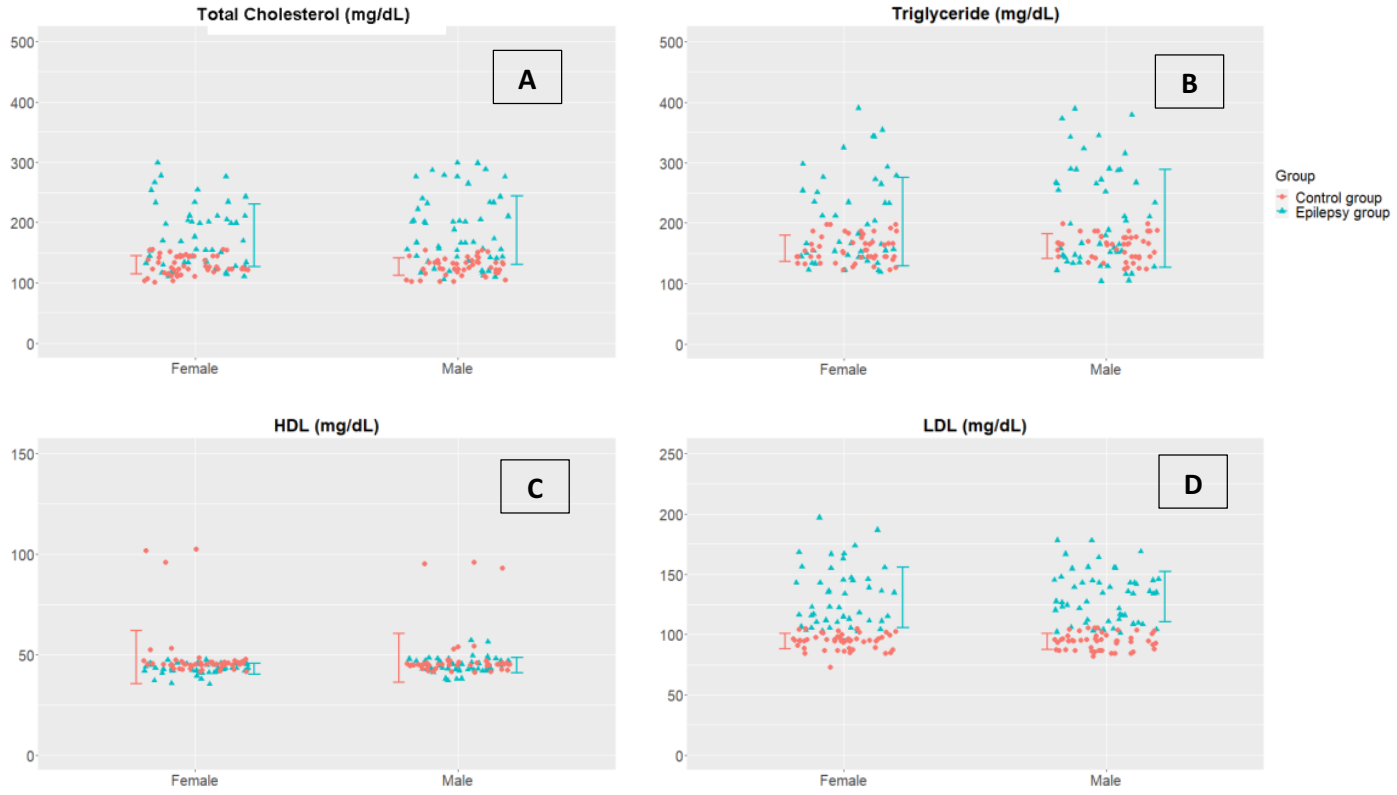

Fig. 1. Comparison of metabolic parameters between epileptic and control groups (female and male). (A) Total cholesterol, (B) Triglyceride, (C) HDL, and (D) LDL. HDL=high-density lipoprotein, LDL=low-density lipoprotein.

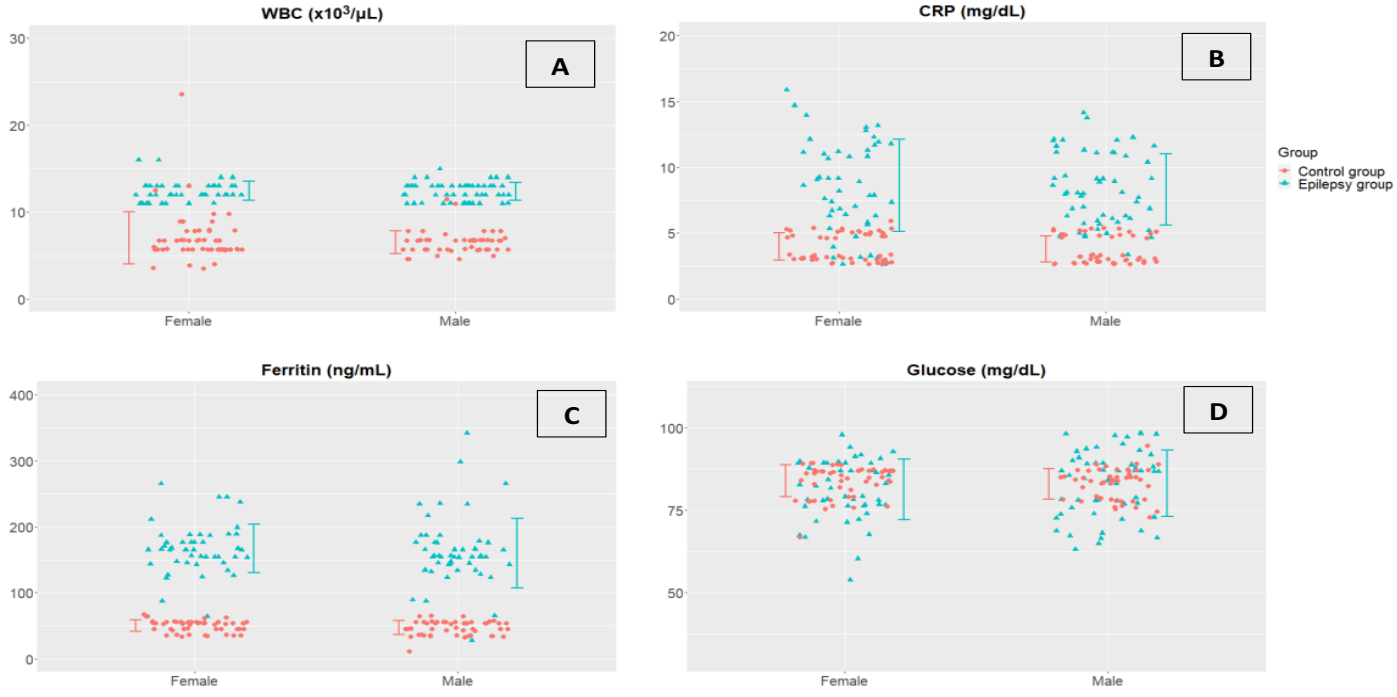

Fig. 2. Comparison of inflammatory markers between the epilepsy and control groups (female and male). (A) WBC, (B) CRP, (C) ferritin, and (D) glucose. WBC=white blood cell, CRP=C-reactive protein.

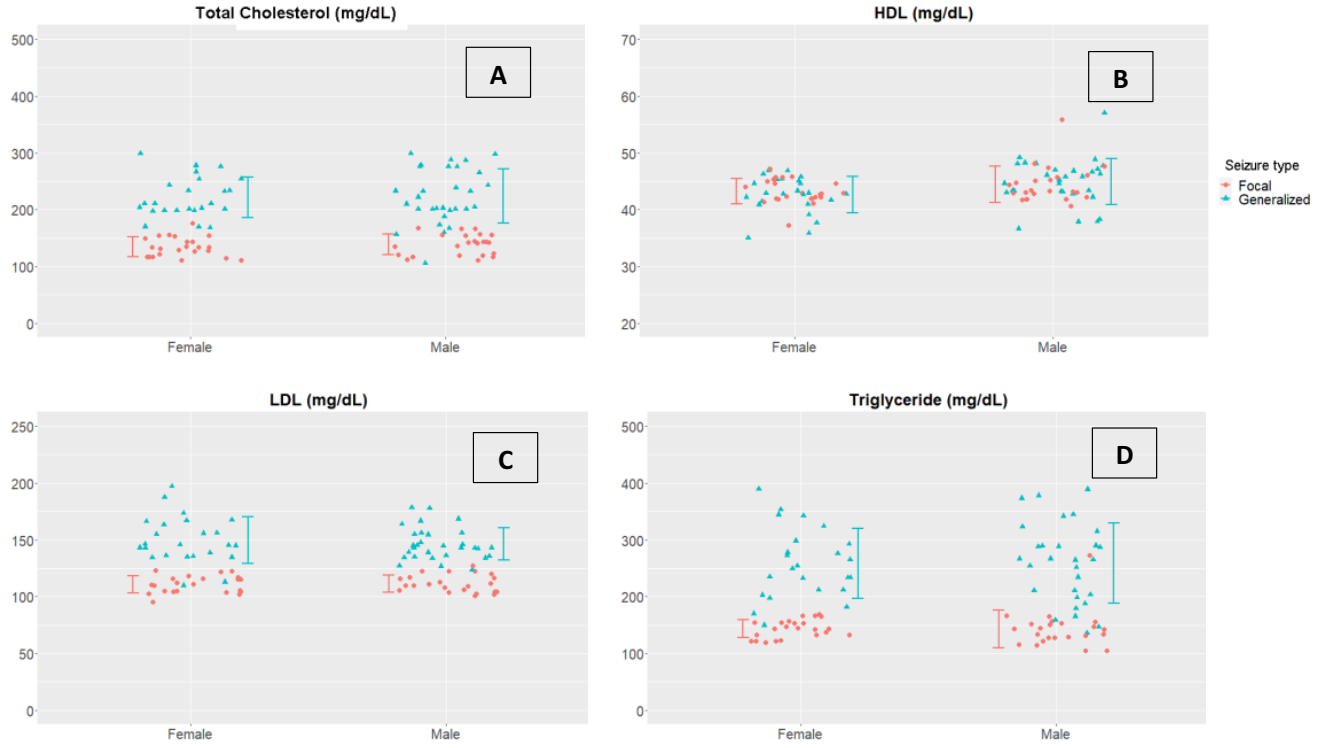

Fig. 3. **The relationship between metabolic markers in patients with generalized and focal seizures. (A) Total cholesterol, (B) HDL, (C) LDL, and (D) Triglyceride. HDL=high-density lipoprotein, LDL=low-density lipoprotein.**

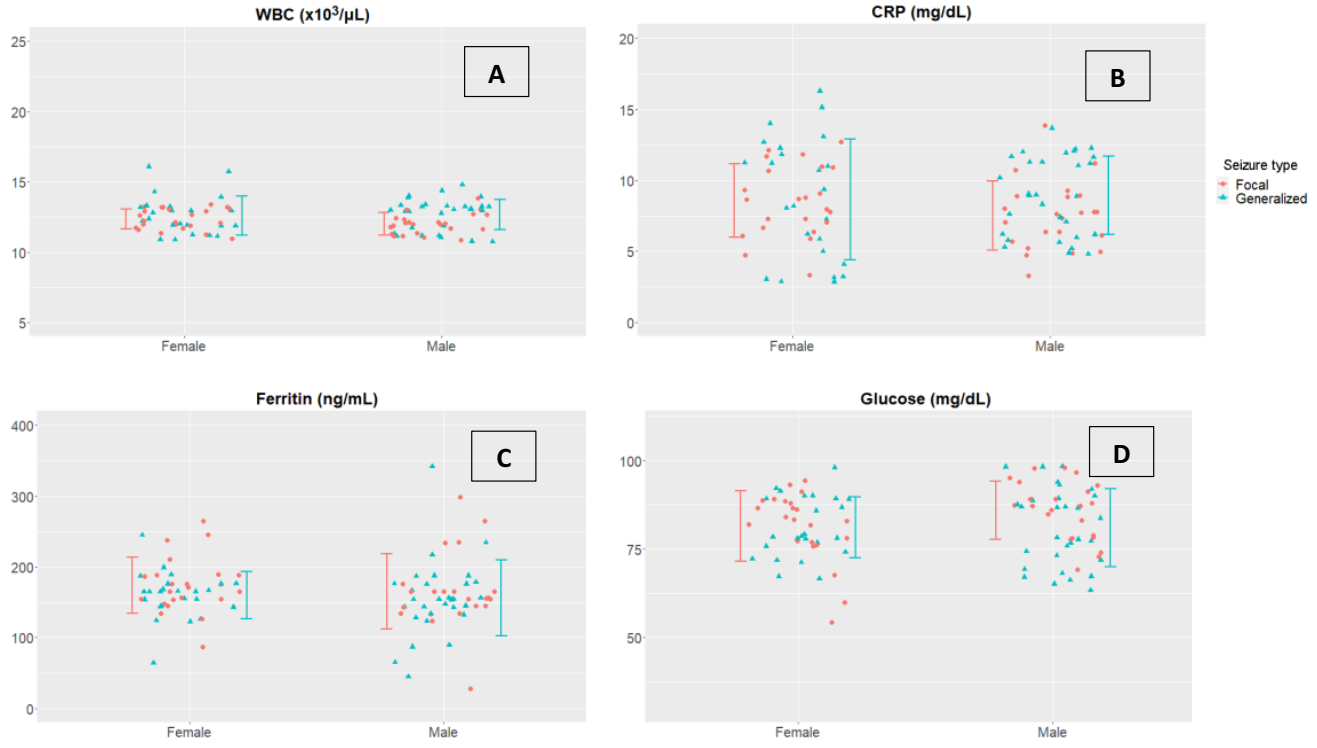

Fig. 4. **Comparison of inflammatory markers. (A) WBC, (B) CRP, (C) ferritin, and glucose (D) in epileptic patients (male and female) with generalized and focal seizures. WBC=white blood cell, CRP=C-reactive protein.**

Table 4. Comparison of lipid profile and inflammatory markers in pediatric epilepsy patients with generalized versus focal seizures

	Seizure type		P value
	Focal (n=46)	Generalized (n=54)	
TC (mg/dL)	136.3±17.7 135 (119-153)	222.9±42.5 211 (199-254)	**<0.001**
TRG (mg/dL)	143.5±25.7 143 (128-154)	258.8±66.3 265 (211-293)	**<0.001**
LDL (mg/dL)	111±7.5 110.5 (105-116)	147.8±17.2 145 (136-156)	**<0.001**
HDL (mg/dL)	43.8±2.8 43 (42-45)	43.9±3.8 44 (42-46)	0.884
WBC ($\times 10^3/\mu L$)	12.2±0.8 12 (12-13)	12.6±1.2 13 (12-13)	**0.031**
CRP (mg/dL)	8±2.5 8 (6-9)	8.8±3.5 9 (6-12)	0.227
Ferritin (ng/mL)	169.4±46.6 165 (145-187)	157.7±45.2 155.5 (143-176)	0.209
Glucose (mg/dL)	83.7±9.3 86 (78-89)	81.1±9.9 78.5 (73-89)	0.183

Data are shown as mean±standard deviation or median (IQR) where appropriate. TC=Total cholesterol, TGL=Triglycerides, LDL=Low-density lipoprotein, HDL=High-density lipoprotein cholesterol, WBC=White blood cell, CRP=C reactive protein

metabolic parameters in the clinical workup of pediatric epilepsy patients. Experimental and clinical studies have shown that major inflammatory mediators, particularly cytokines, are produced during epileptiform activity in the regions of the brain where seizures initiate and spread [6]. Moreover, certain cytokines like Interleukin-1 beta (IL-1β) and Tumor Necrosis Factor-alpha (TNF-α) are also present in the cerebrospinal fluid (CSF), as well as peripheral blood mononuclear cells in the plasma of patients with epilepsy [7]. Conversely, oxidative stress and elevated pro-inflammatory cytokines have been noted in the peripheral blood of patients with epilepsy [8-10]. Our observations vouch for the same, with elevated inflammatory and lipid profiles; these processes can lead to increased lipid peroxidation, protein carbonylation, and even DNA damage [11]. As seizure types vary and are associated with varying patterns of brain activity, and there is evidence that inflammatory responses are more diffusely distributed in generalized than in focal seizures [8], the present study also explored the potential relationships between selected inflammatory and oxidative markers, and glycolipid parameters, and seizure types (focal vs. generalized) within the epilepsy group. There was also a positive correlation between elevated levels of WBC and generalized seizures. No such correlations were seen in patients with focal seizures or in controls. Additionally, metabolic parameters such as TC, TGL, and LDL showed significant correlations with generalized seizures in pediatric epilepsy patients. One of the possible reasons for the observations is that seizures affect the microenvironment of the neuronal tissue, and as a consequence, cytokine [12] as well as Reactive Oxygen Species (ROS) secretion by glial cells may occur [2]. This leads to activation of caspase pathways [13] and damage to DNA [14]. In generalized seizures, these alterations are more extensive in the nervous system than in partial seizures and therefore lead to a higher rise of such parameters in the peripheral circulation.

These findings collectively suggest that these parameters can be valuable indices to assess the severity of damage caused by convulsive seizures because the range was wider in those patients with generalized seizures than in those with partial seizures.

In adult patients, cardiovascular diseases (CVD) and CVD events are known to be associated with epilepsy [15-17]. Population-based studies have shown that mortality among people with epilepsy is elevated. CVD has also been demonstrated to be a significant contributor to premature mortality in this population [18, 19]. Within this framework, neuronal injury that disrupts cerebral cholesterol homeostasis may play a significant role in the pathophysiology of epileptic seizures [20]. In this process, oxidized LDL, through the generation of reactive oxygen species, can induce endothelial dysfunction and promote foam cell formation [21-22]. Our results are consistent with Hermann *et al*. [23], who reported the reduction of HDL in epileptic patients. Although HDL is most notably recognized for its health-promoting effects in keeping endothelial channels healthy by a variety of complicated mechanisms [23], HDL can't so readily be labeled as good cholesterol [24-26]. Both compositional modification in HDL, metabolism, and inflammation may reduce its anti-inflammatory and anti-oxidant activity, as well as functional alteration such that HDL can promote a pro-inflammatory and pro-oxidant impact by facilitating the oxidation of LDL [27-30].

Although patients with generalized seizures generally exhibited higher TC, TGL, and LDL levels, some individuals with this seizure type had normal lipid and inflammatory parameters. These findings suggest that factors such as seizure duration, severity, and the timing of blood sampling may influence the results. Therefore, further research is warranted to clarify these associations.

Taken together, our findings indicate that routine monitoring of lipid and inflammatory markers in pediatric epilepsy patients, particularly those with generalized seizures, may provide valuable diagnostic and prognostic insights. Such biomarkers could support clinical decision-making in addition to traditional EEG and neuroimaging, offering a more comprehensive understanding of disease burden. Future prospective studies with larger, multi-center cohorts are needed to confirm these findings. Combining metabolic and inflammatory profiling with genetic and neuroimaging data may help establish reliable biomarker panels for early differentiation of seizure types and prediction of clinical outcomes.

Limitations

Because this is a retrospective study involving a relatively small group of patients, the results should be interpreted with caution. Additional limitations include the absence of genetic analysis, which prevented evaluation of potential genotype-phenotype correlations, and the inability to assess seizure duration or frequency and medication effects because the cohort consisted of newly diagnosed patients.

CONCLUSION

Pediatric patients with epilepsy, particularly those with generalized seizures, exhibit higher lipid and inflammatory marker levels compared to controls and patients with focal seizures. These findings suggest potential diagnostic and prognostic roles for metabolic and inflammatory markers in seizure classification.

Ethics Approval and Consent to Participate

This study was approved by the University of Health Sciences Gazi Yaşargil Training and Research Hospital Clinical Research Ethics Committee (Decision No: 2023/492; date: 04.08.2023). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Written informed consent was obtained from all participants and their legal guardians.

Data Availability

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

Authors' Contribution

Study Conception: SB; Study Design: SB, HB; Supervision: SB; Funding: N/A; Materials: SB; Data Collection and/or Processing: SB, HB; Statistical

Analysis and/or Data Interpretation: SB, HB; Literature Review: SB; Manuscript Preparation: SB; and Critical Review: SB, HB.

Conflict of Interest

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

Financing

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

Acknowledgments

The authors have no acknowledgments to declare.

Generative Artificial Intelligence Statement

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

Editor's Note

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

REFERENCES

1. Vezzani A, French J, Bartfai T, Baram TZ. The role of inflammation in epilepsy. Nat Rev Neurol. 2011;7(1):31-40. doi: 10.1038/nrneurol.2010.178.
2. Lehtimaki KA, Keranen T, Palmio J, et al. Increased plasma levels of cytokines after seizures in localization-related epilepsy. Acta Neurol Scand. 2007;116(4):226-230. doi: 10.1111/j.1600-0404.2007.00882.x.
3. Pascotini ET, Flores AE, Kegler A, et al. Apoptotic markers and DNA damage are related to the late phase of stroke: Involvement of dyslipidemia and inflammation. Physiol Behav. 2015;151:369-378. doi: 10.1016/j.physbeh.2015.08.005.
4. Specchio N, Wirrell EC, Scheffer IE, et al. International League Against Epilepsy classification and definition of epilepsy syndromes with onset in childhood: Position paper by the ILAE Task Force on Nosology and Definitions. Epilepsia. 2022;63(6):1398-1442. doi: 10.1111/epi.17241.
5. Kegler A, Pascotini ET, Caprara ALF, et al. Relationship between seizure type, metabolic profile, and inflammatory markers in blood samples of patients with epilepsy. Epileptic Disord. 2021 Feb 1;23(1):74-84. doi: 10.1684/epd.2021.1236.
6. Mao LY, Ding J, Peng WF, et al. Interictal interleukin-17A levels are elevated and correlate with seizure severity of epilepsy patients. Epilepsia. 2013;54(9):e142-145. doi: 10.1111/epi.12337.
7. Choi J, Koh S. Role of brain inflammation in epileptogenesis. Yonsei Med J. 2008;49(1):1-18. doi: 10.3349/ymj.2008.49.1.1.
8. Fabene PF, Bramanti P, Constantin G. The emerging role for chemokines in epilepsy. J Neuroimmunol. 2010;224(1-2):22-7. doi: 10.1016/j.jneuroim.2010.05.016.
9. Wang C, Zhai J, Zhou X, Chen Y. Lipid metabolism: Novel approaches for managing idiopathic epilepsy. Neuropeptides. 2024;108:102475. doi: 10.1016/j.npep.2024.102475.
10. Soltani Khaboushan A, Yazdanpanah N, Rezaei N. Neuroinflammation and Proinflammatory Cytokines in Epileptogenesis. Mol Neurobiol. 2022;59(3):1724-1743. doi: 10.1007/s12035-022-02725-6.
11. Rowley S, Patel M. Mitochondrial involvement and oxidative stress in temporal lobe epilepsy. Free Radic Biol Med. 2013;62:121-131. doi: 10.1016/j.freeradbiomed.2013.02.002.
12. Frantseva MV, Velazquez JL, Hwang PA, Carlen PL. Free radical production correlates with cell death in an in vitro model of epilepsy. Eur J Neurosci. 2000;12(4):1431-1439. doi: 10.1046/j.1460-9568.2000.00016.x.
13. Leszek J, Barreto GE, Gąsiorowski K, Koutsouraki E, Ávila-Rodrigues M, Aliev G. Inflammatory Mechanisms and Oxidative Stress as Key Factors Responsible for Progression of Neurodegeneration. CNS Neurol Disord Drug Targets. 2016;15(3):329-336. doi: 10.2174/1871527315666160202125914.
14. Skeldon AM, Faraj M, Saleh M. Caspases and inflammasomes in metabolic inflammation. Immunol Cell Biol. 2014;92(4):304-313. doi: 10.1038/icb.2014.5.
15. Budihardjo I, Oliver H, Lutter M, Luo X, Wang X. Biochemical pathways of caspase activation during apoptosis. Annu Rev Cell Dev Biol. 1999;15:269-290. doi: 10.1146/annurev.cellbio.15.1.269.
16. Gaitatzis A, Carroll K, Majeed A, W Sander J. The epidemiology of the comorbidity of epilepsy in the general population. Epilepsia. 2004;45(12):1613-1622. doi: 10.1111/j.0013-9580.2004.17504.x.
17. Elliott JO, Lu B, Shneker B, Charyton C, Layne Moore J. Comorbidity, health screening, and quality of life among persons with a history of epilepsy. Epilepsy Behav. 2009;14(1):125-129. doi: 10.1016/j.yebeh.2008.10.013.
18. Janszky I, Hallqvist J, Tomson T, Ahlbom A, Mukamal KJ, Ahnve S. Increased risk and worse prognosis of myocardial infarction in patients with prior hospitalization for epilepsy. Brain. 2009;132(Pt 10):27982-804. doi: 10.1093/brain/awp216.
19. Ding D, Wang W, Wu J, et al. Premature mortality in people with epilepsy in rural China: a prospective study. Lancet Neurol. 2006;5(10):823-827. doi: 10.1016/S1474-4422(06)70528-2.
20. Neligan A, Bell GS, Johnson AL, Goodridge DM, Shorvon SD, Sander JW. The long-term risk of premature mortality in people with epilepsy. Brain. 2011;134(Pt 2):388-395. doi: 10.1093/brain/awq378.

21. Adibhatla RM, Hatcher JF. Altered lipid metabolism in brain injury and disorders. Subcell Biochem. 2008;49:241-268. doi: 10.1007/978-1-4020-8831-5_9.
22. Park YM. CD36, a scavenger receptor implicated in atherosclerosis. Exp Mol Med. 2014;46(6):e99. doi: 10.1038/emm.2014.38.
23. Hermann BP, Sager MA, Koscik RL, Young K, Nakamura K. Vascular, inflammatory, and metabolic factors associated with cognition in aging persons with chronic epilepsy. Epilepsia. 2017;58(11):e152-e156. doi: 10.1111/epi.13891.
24. Norata GD, Catapano AL. Molecular mechanisms responsible for the anti-inflammatory and protective effect of HDL on the endothelium. Vasc Health Risk Manag. 2005;1(2):119-129. doi: 10.2147/vhrm.1.2.119.64083.
25. Stredny C, Rotenberg A, Leviton A, Loddenkemper T. Systemic inflammation as a biomarker of seizure propensity and a target for treatment to reduce seizure propensity. Epilepsia Open. 2023;8(1):221-234. doi: 10.1002/epi4.12684.
26. Kobylarek D, Iwanowski P, Lewandowska Z, et al. Advances in the Potential Biomarkers of Epilepsy. Front Neurol. 2019;10:685. doi: 10.3389/fneur.2019.00685.
27. Guo HL, Dong N, Hu YH, et al. Serum HDL-C levels in children with epilepsy: a single-center retrospective study. Front Nutr. 2025;12:1523426. doi: 10.3389/fnut.2025.1523426.
28. Aygun D, Uzun H. Association of Comorbidity and Inflammatory and Nutritional Markers with Epilepsy and Seizure Frequency. Nutrients. 2025;17(11):1847. doi: 10.3390/nu17111847.
29. Guo HL, Wang WJ, Dong N, et al. Integrating metabolomics and lipidomics revealed a decrease in plasma fatty acids but an increase in triglycerides in children with drug-refractory epilepsy. Epilepsia Open. 2023;8(2):466-478. doi: 10.1002/epi4.12712.
30. Chen J, Yu H, Liu H, et al. Causal relationship between immune cells and epilepsy mediated by metabolites analyzed through Mendelian randomization. Sci Rep. 2024;14(1):19644. doi: 10.1038/s41598-024-70370-1.

The European Research Journal 2025;11(5):944-956 Original Article

DOI: https://doi.org/10.18621/eurj.1654844 *Public Health*

The impact of psychological status and quality of life of employees on presenteeism in the health-work relationship after earthquakes in Türkiye

Mehmet Koca[1], Feyza İnceoğlu[2], Serdar Deniz[3]

[1]Depatment of Social Services, Faculty of Health Sciences, Malatya Turgut Özal University, Malatya, Türkiye; [2]Department of Biostatistics, Faculty of Medicine, Malatya Turgut Özal University, Malatya, Türkiye; [3]Department of Public Health, Faculty of Medicine, Malatya Turgut Özal University, Malatya, Türkiye

ABSTRACT

Objectives: This study aims to investigate the workers' general health, psychological status, and quality of life perception on presenteeism after the major earthquakes on Türkiye in the relationship between health and labor.

Methods: This study was conducted at Baykan Denim Company, which manufactures in Malatya, with the participation of 327 employees. Data were collected using the socio-demographic questionnaire form, the Attitude Scale for Determining the Psychological Status of Individuals Exposed to Earthquakes, the General Health Questionnaire-12, the Quality of Life Scale and the Presenteeism Scale.

Results: The findings showed that the Attitude Scale for Determining the Psychological Status of Individuals Exposed to Earthquakes scores had a positive statistically significant effect on Presenteeism (β1=0.443; P=0.001) and General Health (β1=0.495; P=0.001) scores and a negative statistically significant effect on the Quality of Life (β1=-0.145; P=0.001) scores. Presenteeism scores had a positive and statistically significant effect on general health (β1=0.183; P=0.001) and Quality of Life (β1=0.131; P=0.009) scores. Presenteeism scores increased as general health and the Quality of Life scores increased.

Conclusions: This study concluded that individuals were mentally impacted by the earthquake. This impact was seen in individuals' overall health perception and quality of life, leading to elevated presenteeism rates.

Keywords: Earthquake, health, quality of life, presenteeism

On February 6, 2023, two major seismic tremors with sizes of 7.8 Mw and 7.5 Mw happened in progression, nine hours separated, within the Kahramanmaraş area of Türkiye. At slightest 50 thousand individuals lost their lives, and more than 100 thousand individuals were injured in Türkiye after the earthquakes. After the seismic tremor, more than 40 thousand post-quake tremors with sizes of up to 6.7 Mw happened. This earthquake was recorded as the longest, biggest, and most serious seismic tremor in the history of the Republic of Türkiye [1].

Natural events, such as earthquakes, hurricanes, and volcanic eruptions, are uncontrollable natural processes. These natural events may cause significant human and economic losses in settlements, infrastructure elements, and agricultural areas in societies that cannot be adequately prepared for disasters for various reasons [2]. In addition to these losses, earthquakes

Received: March 10, 2025 **Accepted:** June 15, 2025 **Available Online:** June 19, 2025 **Published:** September 4, 2025

How to cite this article: Koca M, İnceoğlu F, Deniz S. The impact of psychological status and quality of life of employees on presenteeism in the health-work relationship after earthquakes in Türkiye. Eur Res J. 2025;11(5):944-956. doi: 10.18621/eurj.1654844

Corresponding author: Mehmet Koca, MD., Phone: +90 422 846 12 55, E-mail: mehmet.koca@ozal.edu.tr

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

can also profoundly affect social life and are disasters that have the potential to cause severe psychological effects on people [3]. Studies show that people are affected in various ways after an earthquake. It was determined that people were psychologically affected by the 8.0 magnitude earthquake in Wenchuan in 2008. There was a significant correlation between variables such as job satisfaction, life satisfaction and personal health perception [4]. A study by Oishi *et al*. found that those whose homes were damaged in the Hanshin-Awaji Earthquake reported lower life satisfaction, more negative emotions, and more health problems than those whose homes were not damaged, even 16 years after the earthquake [5]. Türkan and Hatipoğlu concluded that individuals who experienced the 2023 Türkiye earthquake and participated in voluntary aid activities had higher depression and anxiety scores than those who were not exposed to the earthquake [6].

Individuals physically and psychologically affected by disasters, such as earthquakes, may experience a loss of productivity in their working lives. This loss of productivity can occur in two ways. The first is the employee's physical inability to be at work (absenteeism). The second is the inability to fully devote oneself to work despite being at work. The inability of an individual to fully devote oneself to work despite being at work, that is, being present at work, is expressed in English as the word presenteeism. The word "Presenteeism" is derived from the word "presence." "Presence" as a word means being present, being ready, being there, and appearing there. Based on the meaning of the word, "Presenteeism" is used for the loss of productivity due to the inability of the person to fully devote himself/herself to the work for various reasons although he/she is apparently present [7].

There are various definitions of presenteeism. According to Schultz and Edington, presenteeism is defined as "the reduction in the ability of employees to work due to a physical or health-related illness and is measured by costs related to loss of productivity, work errors, and falling below production standards" [8]. According to Martinez and Ferreira, presenteeism is "the fact that employees are at work despite not being able to work at full efficiency due to illness or health conditions" [9]. In these definitions, the concept of "health" comes to the fore. The World Health Organization (WHO) defined health as "not merely the absence of disease but a state of complete physical, mental and social well-being." After this definition, the concept of quality of life in measuring "well-being" has gained increasing importance in health services research and practice [10].

Although the concept of quality of life is often associated with health, it is actually a multi-dimensional concept. This concept is affected by many elements, such as health, education, the economy, and the environment. In this context, quality of life is defined as a state of well-being that includes material and spiritual conditions directly affecting an individual's life [11]. Health-workforce interaction is among the crucial factors in the development of societies.

Health is one of the fundamental elements of economic development and has an important place in reducing poverty and inequalities. Spending on health services to improve the health of individuals and society develops human capital and contributes to economic growth. With economic growth, human capital investments increase, which leads to chain growth [12]. Individuals and societies need to be healthy. Destructive earthquakes have a significant place in human life, and determining the stress experienced after an earthquake, the perception of quality of life, and its impact on working life are critical in developing socio-economic policies for the effective management of the post-disaster period. To our knowledge, no study was found examining the effects of destructive earthquakes on the general and mental health of individuals, the quality of life, and the presenteeism of workers in their working lives, and conducting a study on this subject is important in terms of filling an important gap in the literature.

This study aims to investigate the workers' general health, psychological status, and quality of life perception on presenteeism after the major earthquakes on Türkiye in the relationship between health and labor.

METHODS

Type and Hypotheses of this Research

This study used a relational screening model. The main hypothesis of this study is stated below:

➤H1: The Psychological Status scores of those exposed to the earthquake influence General Health, WHOQOL, and presenteeism scores.

➤H2: The General Health Questionnaire scores

mediate the association between the psychological state of individuals exposed to the earthquake and presenteeism.

➤H3: WHOQOL ratings mediate the link between the psychological state of individuals exposed to an earthquake and presenteeism.

Place and Time of this Research

The data were gathered online and in person from Baykan Denim Company employees in Malatya between February 19, 2024, and May 10, 2024. The company is among the foremost enterprises in Malatya's textile industry, employing roughly 1,000 individuals.

Sample Selection and Number of Samples

Although there is no clear expression for the Structural Equation Model (SEM), it is reported that studies are using 250-500 sample sizes [13]. Kline [14] suggests that the sample size should be 200 or more in analyses conducted with SEM. In line with these views, n=327 participants were included in the present study. Participants were chosen by voluntary sampling and snowball sampling techniques, both of which are non-probability sampling approaches.

Data Collection Tools

Personal Information Form

A 10-question questionnaire was applied to determine the socio-demographic characteristics of the participants in this study.

General Health Questionnaire

This is a scale developed by Goldberg [15] in the 1970s. After the 60-question form, short forms with 30, 28, and 12 questions were also developed and found to be equally reliable. The General Health Questionnaire-12 (GHQ-12) comprises 12 items. The Cronbach's alpha coefficient for internal consistency of the scale was 0.78 [15]. It has been adapted to Turkish by Kılıç [16]. The scale is designed as a 4-point Likert type and is scored as 0-1-2-3 or, as advised in the GSA handbook, as 0-0-1-1. Elevated scores on this scale indicate a heightened prevalence of psychological issues, specifically anxiety and sadness [16].

Attitude Scale to Determine the Psychological States of Individuals Exposed to an Earthquake (ASDPSIEE)

It was created by Filiz *et al*. [17] to assess the psychological conditions of individuals affected by an earthquake. The scale was developed using a 5-point Likert format and comprises six dimensions and 41 statements categorized as "Detachment from Life, Social Health, Spiritual Change, Trauma Anxiety, Maturation, and Avoidance." The scale lacks a reverse expression. A high score signifies that the pertinent dimension is regarded at an elevated degree. The Cronbach's alpha coefficient for the scale created by Filiz *et al*. [17] was 0.96.

Quality of Life Scale (EUROHIS WHOQOL-8.Tr)

The “Quality of Life Scale” is a general-purpose Health Quality of Life (HQOL) scale created by selecting specific questions from the EUROHIS-QOL.8 (WHOQOL- 8) WHOQOL-Bref scale produced from the World Health Organization Quality of Life Scale (WHOQOL) [18]. The scale consists of 8 questions. It was prepared in a 5-point Likert type. As the scores obtained from the scale increase, the quality of life also improves [19, 20]. The Turkish validity and reliability assessment of the scale was performed by Eser *et al*. [21]. Two questions of the WHOQOL scale are aimed at determining general health and general quality of life, and the remaining six questions are aimed at determining physical, mental, social, and environmental dimensions [21].

Presenteeism Scale

The "Stanford Presenteeism Scale (SP 6)," created by Koopman *et al*. [22] and comprising six items, was utilized. The scale consists of two sub-dimensions: distraction avoidance (items 1, 3, and 4) and task completion (items 2, 5, and 6). The reliability investigation yielded a Cronbach’s alpha value of 0.80 [22]. A study on the validity and reliability of the Turkish adaptation of the scale has been undertaken by several authors [23-25]. The Stanford Presenteeism Scale is a 5-point Likert-type scale. While items 1, 3, and 4 in the scale are scored directly, questions 2, 5, and 6 are scored reversely. The total score varies from 6 to 30. High scores indicate a high level of presenteeism.

Ethical Aspects of this Research

Ethics committee approval was received for the study from Malatya Turgut Özal University Non-interventional Clinical Research Ethics Committee with

the decision dated 15.02.2024 and numbered E-30785963-020-208770. This study was done in conformity with the Declaration of Helsinki, and consent was acquired from the participants.

Statistical Analysis

Data analysis in the study was performed using SPSS (Statistical Program in Social Sciences) 28 and AMOS 24 statistical software programs. Kolmogorov-Smirnov technique was used to control normal distribution. The significance level (p) was taken as 0.05 for comparison tests. Mann-Whitney U test with Bonferroni correction was used for independent two-group variables, and the Kruskal-Wallis test was used for in-

Table 1. Demographic information

Variable		Data
Age (years)		34.86±8.24 (18-60)
Gender, n (%)	Female	116 (35.5)
	Male	211 (64.5)
Age group, n (%)	18-30 years old	114 (34.9)
	31-40 years old	122 (37.3)
	Ages 41 and above	91 (27.8)
Educational status, n (%)	Primary education	123 (37.6)
	High school	152 (46.5)
	University and above	52 (15.9)
Marital status, n (%)	Married	187 (57.2)
	Single	140 (42.8)
Having children, n (%)	Yes	182 (55.7)
	No	145 (44.3)
Disabled person responsible for caring for family, n (%)	Yes	64 (19.6)
	No	263 (80.4)
Working hours in the factory, n (%)	Less than 1 year	36 (11.0)
	1-3 years	106 (32.4)
	4-10 years	163 (49.8)
	11-20 years	22 (6.7)
Residence during an earthquake, n (%)	Homeowner	178 (54.4)
	Tenant	149 (45.6)
Condition of the residence during the earthquake, n (%)	Undamaged	56 (17.1)
	Slightly damaged	163 (49.8)
	Medium damaged	22 (6.7)
	Severely damaged	54 (16.5)
	Ruined	32 (9.8)
Current place of stay, n (%)	In my current home	209 (63.9)
	In the container	77 (23.5)
	With my relatives	21 (6.4)
	Other	20 (6.1)

Data are shown as mean±standard deviation (minimum-maximum) or n (%)

dependent multi-group variables. Numbers and percentages were preferred as descriptive values for categorical data, and mean and standard deviation were preferred for quantitative data.

For the multiple normal distribution control, the "Observations farthest from the centroid (Mahalonobis Distance) Menu" in the AMOS program was checked, and the skewness value was calculated as 2.998. Since this value is less than 8, it was assumed to provide a multivariate normal distribution [26]. Since the normal distribution was provided, the Pearson correlation coefficient was used to compare quantitative variables. Structural Equation Modelling (SEM), a mediation analysis method based on the bootstrap method, was preferred to examine the mediation effect between the scales.

The SEM method has been reported to be more reliable than the classical method and the results were obtained with the Sobel test. We reloaded five thousand samples to apply the bootstrap method. The lower and upper limits of the 95% confidence interval (CI) found by the bootstrap method were interpreted by seeing if they included zero (0) values. The model fit indices were used to see if the models were significant [27].

RESULTS

Demographic information about the participants included in this study is given in Table 1, and descriptive statistics of scale scores are in Table 2 below.

Of the participants, 64.5% (n=211) were male, 37.3% (n=122) were between the ages of 31-40, 46.5% (n=152) were high school graduates, 57.2% (n=187) were married, 55.7% (n=182) had children, and 80.4% (n=263) had no dependent disabled person in the family, 49.8% (n=163) had been working in the factory for 4-10 years, 54.4% (n=178) stated that the house they were residing in at the time of the earthquake belonged to them and 63.9% (n=209) stated that they were currently residing in their own house. The average age of the study participants was calculated as 34.86.

Cronbach's α values of the scales used were higher than 0.70, indicating high reliability [28]. In our study, Cronbach's α values of the scales were between 0.78 and 0.86 (Table 2). A comparison of scale scores according to demographic variables is presented in Table 3.

A statistical difference was found in the Presenteeism Scale scores for gender (P<0.05), but no difference was found in the Psychological Status of Individuals Exposed to the Earthquake, General Health, and WHOQOL Scales (P>0.05). No difference was found in Psychological Status, General Health, WHOQOL and Presenteeism Scales of Individuals Exposed to Earthquake according to age, educational status and marital status (P>0.05).

A statistical difference was found between those with and without children, those with and without disabled dependents in the family, and those who owned or did not own the house where they lived during the earthquake in the scale scores of Psychological Status of Individuals Exposed to Earthquake (P<0.05), but no difference was found in Presenteeism, General Health and WHOQOL Scales (P>0.05).

A statistical difference was found in the Psychological Status of Individuals Exposed to an Earthquake scale scores according to the conditions of the residence (P<0.05), but no difference was found in the Presenteeism, General Health, WHOQOL Scales (P>0.05). According to the conditions of the residence, there was a difference between houses with severe damage and undamaged houses (P=0.001). There was a difference between houses with moderate damage and undamaged houses (P=0.002). There was a difference between houses with moderate damage and

Table 2. Descriptive statistics of scale scores

Scale	Mean±SD	(Min-Max)	Cronbach's
ASDPSIEE	130.39±43.27	41-205	0.78
General health	14.5±8.13	0-36	0.86
WHOQOL	24.07±7.5	8-40	0.82
Presenteeism	16.9±6.19	6-30	0.82

SD=standart deviation, Min-Max=minimum-maximum

Table 3. Comparison of scale scores according to demographic variables

Variables	Groups	ASDPSIEE		General health		WHOQOL 8		Presenteeism	
		Mean±SD	Mean (Min-Max)	Mean±SD	Mean (Min-Max)	Mean±SD	Mean (Min-Max)	Mean±SD	Median (Min-Max)
Gender	**Female**	133.17±40.26	133.5 (41-205)	14.07±8.87	12 (0-36)	24.89±6.59	25 (8-40)	17.92±5.97	18 (6-30)
	Male	129.29±44.56	128 (41-205)	14.76±7.72	12.5 (0-35)	23.62±7.94	24(8-40)	16.38±6.24	18 (6-30)
Mann Whitney		11543.500		11259.000		11094.500		**10095.000**	
P value		0.396		0.231		0.182		**0.009***	
Age (years)	**18-30 years old**	129.46±45.81	127.5 (41-205)	13.81±7.99	12 (0-36)	24.67±7.74	25 (8-40)	17.08±6.56	18 (6-30)
	31-40 years old	127.67±40.86	129 (41-205)	14.83±8.39	12 (0-36)	23.15±8.39	24 (8-40)	17.21±5.84	18 (6-30)
	41 years and older	136.18±42.28	147 (41-205)	14.98±8.02	14 (0-34)	24.55±5.69	24 (12-40)	16.36±6.17	17 (6-30)
Kruskal Wallis		2.250		1.805		2.731		1.166	
P Value		0.325		0.406		0.255		0.558	
Educational status	**Primary education**	133.4±41.92	138 (41-205)	14.42±7.94	12 (0-35)	23.5±7.55	24 (8-40)	17.14±6.72	18 (6-30)
	High school	132.22±41.49	130 (41-205)	14.26±7.83	13 (0-35)	24.94±7.15	25 (8-40)	16.74±5.96	18 (6-30)
	University graduate and above	119.71±48.98	123.5 (43-205)	15.46±9.49	12.5 (2-36)	22.88±8.22	24 (8-39)	16.96±5.58	17 (6-30)
Kruskal Wallis		3.202		0.105		3.983		0.360	
P value		0.202		0.949		0.136		0.835	
Marital status	**Married**	131.95±42.68	131 (41-205)	14.59±8.03	13 (0-35)	24.27±7.41	25 (8-40)	17.06±6.18	18 (6-30)
	Single	128.95±43.64	129 (41-205)	14.4±8.31	12 (0-36)	23.81±7.65	24 (8-40)	16.75±6.2	18 (6-30)
Mann Whitney		12215.000		12573.500		12617.500		12510.000	
Sig. (p)		0.301		0.541		0.652		0.492	
Having children	**Yes**	134.16±41.85	140 (41-205)	14.77±8.14	13 (0-35)	24.34±7.45	24 (8-40)	17.09±6.38	18 (6-30)
	No	126.26±44.28	123.5 (41-205)	14.18±8.15	12 (0-36)	23.73±7.58	25 (8-40)	16.72±5.94	18 (6-30)
Mann Whitney		11475.500		12469.000		12988.500		12516.500	
P value		**0.043***		0.392		0.891		0.423	
Disabled person responsible for caring for family	**Yes**	142.73±41.23	152 (52-205)	16.22±9.63	15 (0-36)	24.52±8.47	26 (8-40)	17.83±5.75	18 (6-30)
	No	127.78±43.06	126 (41-205)	14.1±7.71	12 (0-36)	23.96±7.27	24 (8-40)	16.71±6.27	18 (6-30)
Mann Whitney		6853.000		7381.000		7793.500		7610.000	
Sig. (p)		**0.021***		0.127		0.464		0.233	

Table 3 contunied. Comparison of scale scores according to demographic variables

Variables	Groups	ASDPSIEE		General health		WHOQOL 8		Presenteeism	
		Mean±SD	Mean (Min-Max)	Mean±SD	Mean (Min-Max)	Mean±SD	Mean (Min-Max)	Mean±SD	Median (Min-Max)
Working hours in the factory	**Less than 1 year**	130.11±43.16	127.5 (53-195)	13.75±7.1	11 (4-29)	24.5±7.19	24 (11-36)	17.25±4.99	18 (6-30)
	1-3 years	130.42±46.95	129 (41-205)	15.21±7.83	14 (0-36)	24.18±7.77	25 (8-40)	16.86±6.41	18 (6-30)
	4-10 years	127.54±40.84	129 (41-205)	13.51±8.16	12 (0-36)	24.3±7.55	25 (8-40)	16.73±6.33	17 (6-30)
	11-20 years	155.86±32.16	159 (87-189)	19.82±9.14	19 (2-34)	21.18±6.08	21 (9-31)	18.18±5.9	18 (6-29)
Kruskal Wallis		9.473		14.740		3.906		1.530	
P value		**0.024***		**0.002***		0.272		0.675	
Residence during an earthquake	**Homeowner**	126.58±41.68	126 (41-205)	14.81±8.44	12(0-36)	23.86±7.34	24 (8-40)	16.9±6.03	18 (6-30)
	Tenant	135.52±44.28	134 (41-205)	14.15±7.78	12(0-35)	24.32±7.72	25 (8-40)	16.96±6.38	18 (6-30)
Mann Whitney		11557.500		12779.000		12460.500		13250.000	
P value		**0.045***		0.571		0.391		0.990	
Condition of the residence during the earthquake	**Undamaged**	122.73±43.34	119 (41-193)	12.04±7.27	12(0-30)	25.89±8.13	25 (8-40)	16.63±6.61	18 (6-30)
	Slightly damaged	125.45±41.97	126 (41-205)	15.23±8.17	12(0-35)	24.09±7.46	24 (8-40)	17.09±6.2	18 (6-30)
	Medium damaged	142.27±60.23	131 (47-205)	13.09±7.06	13(2-25)	24.23±7.54	25 (12-40)	17.68±7.85	17 (6-30)
	Severely damaged	142.19±38.9	146 (52-205)	15.11±9.26	14(0-36)	22.35±7.72	23 (8-40)	16.74±5.68	18 (6-30)
	Ruined	143.56±33.93	152 (85-198)	15.16±7.58	14.5(5-29)	23.59±5.68	24 (11-34)	16.44±5.1	18 (6-24)
Kruskal Wallis		11.119		5.416		6.075		0.177	
P value		**0.025***		0.247		0.194		0.996	
Current place of stay	**In my current home**	125.79±42.41	125 (41-205)	14.04±7.67	12(0-35)	24.24±7.45	24 (8-40)	16.8±6.29	18 (6-30)
	In the container	143.94±41.87	149 (48-205)	14.84±7.79	13(3-36)	23.71±7.21	24 (8-40)	17.62±5.93	18 (6-30)
	With my relatives	141.9±42.88	134 (82-198)	20.24±9.91	19(6-36)	20.14±6.46	21 (8-30)	16.52±4.73	17 (6-30)
	Other	118.55±43.58	123.5 (41-185)	12.1±10.09	12.5 (0-35)	27.8±8.56	28 (11-40)	16±7.5	16 (6-30)
Kruskal Wallis		11.060		8.903		10.418		1.938	
P value		**0.011***		**0.031***		**0.015***		0.585	

Data are shown as mean±standard deviationor median (minimum-maximum).

*P<0.05; There is a statistical difference between the groups.

houses with severe damage (P=0.004).

A statistical difference was found in the Psychological Status and General Health Scale scores of individuals exposed to the earthquake according to their working hours in the factory (P<0.05), but no difference was found in the Presenteeism and WHOQOL Scales (P>0.05).

According to the working hours in the factory, in the Psychological Status of Individuals Exposed to the Earthquake Scale scores, there was a difference between the 1-3 year and 11-20 year groups (P=0.008). There was a difference between the 4-10 year and 11-20 year groups (P=0.002). In the General Health Questionnaire Scale scores, there was a difference between the 4-10 year and 11-20 year groups (P=0.001).

A statistical difference was found in the Psychological Status of Individuals Exposed to Earthquake, General Health, and WHOQOL Scale scores among individuals according to their current place of residence (P<0.05), but no difference was found in the Presenteeism Scales (P>0.05).

In the scale scores of the Psychological Status of Individuals Exposed to the Earthquake, there was a difference between those who stayed in their current homes and those who stayed in containers (P=0.004). In the General Health questionnaire scale scores, there was a difference between those who stayed in their current homes and those who stayed in containers (P=0.006).

In the WHOQOL Scale scores, there was a difference between those who stayed with their relatives and those who stayed in other places (P=0.004).

Results of SEM Analysis

SEM was established with observed variables to examine the relationship between the Psychological Status, General Health, WHOQOL and Presenteeism Scales of Individuals Exposed to Earthquake. The established model is given in Fig. 1. The coefficients of the model are given in Table 4.

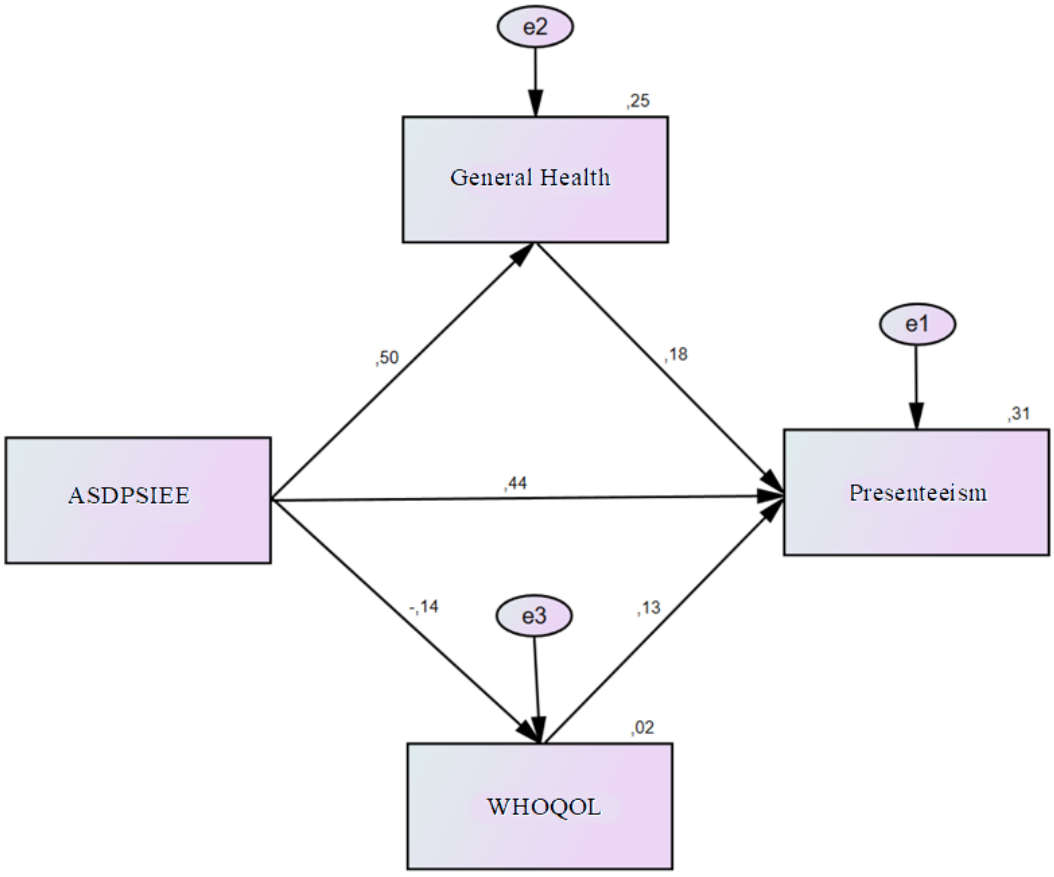

Fig. 1. Model Diagram of the Relationship between the Psychological Status, General Health, WHOQOL and Presenteeism Scales of Individuals Exposed to the Earthquake.

In structural equation modeling analyses where more than one regression model is analyzed simultaneously, a variable can be both a dependent and independent variable simultaneously. Structural equation modeling analyses provide researchers with more interpretation opportunities [29].

ASDPSIEE scores had a positive statistically significant effect on Presenteeism (β1= 0.443; P=0.001) and General Health (β1=0.495; P=0.001) scores and a negative statistically significant effect on WHOQOL (β1=-0.145; P=0.001) scores. As ASDPSIEE scores increase, Presenteeism and General Health scores will increase, and WHOQOL scores will decrease. Presenteeism scores had a positive statistically significant effect on General Health (β1=0.183; P=0.001) and WHOQOL (β1= 0.131; P=0.009) scores. As General Health and WHOQOL scores increase, Presenteeism scores will increase.

ASDPSIEE scores explained 3% of WHOQOL scores (R2=0.03) and 25% of General Health scores (R2=0.25). The effect of ASDPSIEE scores on General Health was higher than its effect on WHOQOL. ASDPSIEE, WHOQOL and General Health Scores together explained 31% (R2=0.31) of Presenteeism scores. 65.3% of hyperarousal scores and 58.8% of avoidance scores were explained by reliving scores. 55.1% of negative change scores were explained by avoidance and reliving scores.

In structural equation modeling, which use many fit indices rather than a singular fit index, the accuracy of the established model is evaluated based on the factors identified in the research. Within the newly made model, according to the analysis results, the goodness of fit index value χ2/df (Chi-Square Goodness of Fit; χ2, df; degree of freedom) was found to be 4.372. The RMSEA (Root Mean Square Error of Approximation) value of 0.075 (RMSEA<0.80), which is the index showing the adequacy of the sample size, shows that the sample size is sufficient for the model used. The GFI (Goodness of Fit Index) value was 0.939, CFI (Comparative Fit Index) was 0.926, IFI (Incremental Fit Index) was 0.927, and NFI (Normed Fit Index) was 0.924. In terms of fit indices, the model fit was seen to be very good [29, 30].

In the mediation model established for general health and WHOQOL, the new approach Bootstrap results found that the indirect effects of ASDPSIEE scores on presenteeism were statistically significant (β= 0.072, CI [0.019-0.129]). It was observed that the Bootstrap lower confidence interval (0.019) and upper confidence interval (0.129) obtained using the percentage method did not include the value zero (0). The mediator model was statistically significant [31, 32]. The indirect effect of general health and WHOQOL in explaining the indirect effect of ASDPSIEE scores on Presenteeism was 0.72.

DISCUSSION

According to the International Disaster Database (EM-DAT) developed by the Disasters Epidemiology Research Center, in 2021, 432 catastrophic events worldwide caused 10,492 deaths and economic losses of approximately 252.1 billion USD [33]. In Türkiye, the approximate cost of the Kahramanmaraş earthquakes in 2023 is estimated to be 103.6 billion dollars [34].

The deterioration of the health of individuals negatively affects all areas together with economic systems. Because the capacity of the labor force is one of the essential elements of production, doing business decreases when health deteriorates [35]. While it is

Table 4. Descriptive values of model coefficients

Dependent Variable	Independent variable	β_1	β_2	P value	R^2
Presenteeism	**ASDPSIEE**	0.443	0.064	**0.001***	0.31
	WHOQOL_8	0.131	0.108	**0.009***	
	General health	0.183	0.14	**0.001***	
WHOQOL_8	**ASDPSIEE**	-0.145	-0.025	**0.008***	0.03
General health	**ASDPSIEE**	0.495	0.093	**0.001***	0.25

β_1=Standardized regression coefficients, β_2=Unstandardized regression coefficients, R^2=Explanatory coefficients
*P<0.05; t test result for the significance of the regression coefficients

easier to measure the loss of productivity when individuals are not directly at work (absenteeism), it is more challenging to measure the loss of productivity due to presenteeism. However, according to calculations, it is stated that the loss due to presenteeism is greater than the loss due to absenteeism [7, 36, 37]. Stewart *et al*. [38] reported that the loss of productive time due to widespread pain in a study conducted among active workers in the United States reached approximately $61.2 billion per year and found that the vast majority of the lost productive time (76.6%) was due to poor performance while at work, not absenteeism [38]. Hemp [39] reported that 63% of the financial loss was due to presenteeism, which cost $311.8 million.

In this study, a statistically significant difference was found in the presenteeism scale scores for gender. There are different results in the studies conducted in the literature [40-43]. Bulan and Söyük [44] examined 46 theses on presenteeism in our country between 2007 and 2022 and determined that women generally have higher levels of presenteeism than men. They stated that women are at higher risk of stress and depression than men and that this increases presenteeism behavior in the workplace [22, 45]. No relationship was found between presenteeism and age, marital status, education level, and length of service at the workplace. This result is consistent with the results of the study conducted by Yılmaz [40] in 2019 with employees of a textile company in Edirne [40].

ASDPSIEE scores have a positive statistically significant effect on Presenteeism (β1= 0.443; P=0.001) and General Health (β1=0.495; P=0.001) scores and a negative statistically significant effect on WHOQOL (β1=-0.145; P=0.001) scores. As ASDPSIEE scores increase, Presenteeism and General Health scores will increase, and WHOQOL scores will decrease. The first hypothesis of this study, "H1: The Psychological Status scores of those exposed to the earthquake influence General Health, WHOQOL, and Presenteeism scores." hypothesis was accepted.

Presenteeism scores have a positive statistically significant effect on General Health (β1=0.183; P=0.001) and WHOQOL (β1= 0.131; P=0.009) scores. ASDPSIEE scores explain 3% of WHOQOL scores (R2=0.03) and 25% of General Health scores (R2=0.25). This result has determined that individuals' psychology is affected after major traumas, affecting their perception of health and life and increasing presenteeism in working life. Howard *et al*. [46] found in their study that occupational and health factors mediate stress, and stress contributes significantly to presenteeism at work.

In a study conducted by Li *et al*. [47] on Chinese nurses, it was determined that health was significantly associated with absenteeism and loss of productivity in nurses and had a mediating effect. It was determined that health played a fully mediating role between nurses' presence at work and loss of productivity, and the indirect effect explained 36% of the total effect [47]. This study determined that ASDPSIEE, WHOQOL and General Health scores together explained 31% (R2=0.31) of Presenteeism scores.

The mediator model established in this study was statistically significant. In explaining the indirect effect of ASDPSIEE scores on Presenteeism, the indirect effect of General Health and WHOQOL was calculated as 0.72. In the light of these data, the research hypotheses “H2: The General Health Questionnaire scores mediate the association between the psychological state of individuals exposed to the earthquake and presenteeism.” and “H3: WHOQOL ratings mediate the link between the psychological state of individuals exposed to an earthquake and presenteeism.” were accepted. It is stated that those who evaluate their own health as poor may be more likely to experience presenteeism [48]. Magalhaes *et al*. [49] found in their study that quality of life is significantly correlated with presenteeism. Aronsson *et al*. [50] also determined in their study that health and motivation are correlated with presenteeism and absenteeism. Studies show that people are directly affected by events that directly affect their lives, and this influence is reflected in their daily lives.

Limitations

he study results are limited to the factory where the study was conducted. There may be limitations arising from the sampling method of the study.

CONCLUSION

This study concluded that individuals were mentally impacted by the earthquake. This effect was evident in individuals' overall health perception and quality of life,

leading to elevated presenteeism rates. It was established that general health perception scores and quality of life scores mediated the relationship between the psychological attitudes of those affected by the earthquake and presenteeism. The combined ASDPSIEE, WHOQOL, and General Health scores accounted for 31% of the variance in Presenteeism scores.

Protecting the health and work-life balance of the workforce is essential for the advancement of nations and enterprises, and it is vital for persons impacted by severe disasters, such as earthquakes, to expedite their recovery from their circumstances. The outcomes of this study suggest that implementing economic and psychological policies at both micro and macro levels would be advantageous for individuals.

Ethical Statement

This study was approved by the Malatya Turgut Özal University Non-interventional Clinical Research Ethics Committee (Decision no.: E-30785963-020-208770, date: 15.02.2024). This study was done in conformity with the Declaration of Helsinki, and consent was acquired from the participants.

Data Availability

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

Authors' Contribution

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

Conflict of interest

The authors disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The authors disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgements

We would like to thank Malatya Baykan Denim Human Resources Manager Mr. Mustafa Bakan, Baykan Denim senior management and employees for their interest and contribution during the planning and implementation of the study.

Generative Artificial Intelligence Statement

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

Editor's note

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

REFERENCES

1. 2023 Kahramanmaraş depremlerindeki kayıplar. Available at: https://tr.wikipedia.org/wiki/2023_Kahramanmara%C5%9F_depremlerindeki_kay%C4%B1plar. Accessed September 01, 2024.
2. Chaudhary MT, Awais P. Natural disasters - origins, impacts, management. Encyclopedia. 2021;1(4):1101-1131. doi: 10.3390/encyclopedia1040084.
3. Nakajima Ş. [Post-Earthqake Psychology]. Okmeydanı Tıp Derg. 2012;28(Suppl 2):150-155. doi: 10.5222/otd.supp2.2012.150. [Article in Turkish].
4. Xu J, Wu Z. One-year follow-up analysis of cognitive and psychological consequences among survivors of the Wenchuan earthquake. Int J Psychol. 2011;46(2):144-152. doi: 10.1080/00207594.2010.529908.
5. Oishi S, Kimura R, Hayashi H, et al. Psychological adaptation to the Great Hanshin-Awaji Earthquake of 1995: 16 years later victims still report lower levels of subjective well-being. J Res Pers. 2015;55:84-90. doi: 10.1016/j.jrp.2015.02.001.
6. Türkkan T, Hatipoğlu E. Factors affecting the mental health of youth after the great Kahramanmaras earthquake disaster. J Soc Serv Res. 2025;51(2): 335-347. doi: 10.1080/01488376.2024.2352158.
7. Çiftçi B. [The Problem of Presenteeısm And Solutions For Busıness]. Çalışma ve Toplum. 2010;1(24):153-174. [Article in Turkish].
8. Schultz AB, Edington DW. Employee health and presenteeism: a systematic review. J Occup Rehabil. 2007;17(3):547-579. doi:

10.1007/s10926-007-9096-x.
9. Martinez LF, Ferreira AI. Sick at work: presenteeism among nurses in a Portuguese public hospital. Stress Health. 2012;28(4):297-304. doi: 10.1002/smi.1432.
10. Testa MA, Simonson DC. Assessment of quality-of-life outcomes. N Engl J Med. 1996;334(13):835-840. doi: 10.1056/NEJM199603283341306.
11. Uygun Şahin Ö. [Quality of life and global climate change]. J Awareness. 2021;6(3):147-154. doi: 10.26809/joa.6.3.06. [Article in Turkish]
12. Akıncı A, Tuncer G. [The Relationship Between Health Expenditures and Economic Growth in Turkey]. Sayıştay Derg. 2016;102:47-61. [Article in Turkish]
13. Schumacker RE, Lomax RG. A Beginner's Guide to Structural Equation Modeling. 3rd ed., Taylor and Francis Group, LLC: New York; 2010.
14. Kline RB. Principles and Practice of Structural Equation Modeling. 3rd ed., The Guilford Press; 2011:209-342.
15. Goldberg DP. Manual of the General Health Questionnaire. Windsor: NEFR Publishing; 1978.
16. Kılıç C. Genel Sağlık Anketi: Geçerlik ve güvenirlik çalışması. Turk Psikiyatri Derg. 1996;7(1):3-9. [Article in Turkish].
17. Filiz M, Karagöz Y, Karaşin Y. [Scale Development Study to Determine the Psychological Effects Of Individuals Exposed to Earthquake]. Dicle Univ Sos Bil Enst Derg. 2023;33:236-250. doi: 10.15182/diclesosbed.1263545. [Article in Turkish]
18. The World Health Organization Quality of Life Assessment (WHOQOL): development and general psychometric properties. Soc Sci Med. 1998;46(12):1569-85. doi: 10.1016/s0277-9536(98)00009-4.
19. Power M. Development of a common instrument for quality of life. In: Nosikov A, Gudex C, eds. EUROHIS: Developing Common Instruments for Health Surveys. IOS Press; 2003: pp.145-163.
20. Schmidt S, Muhlan H, Power M. The EUROHIS-QOL 8-item index: Psychometric results of a cross-cultural field study. Eur J Public Health. 2006;6(4):420-428. doi: 10.1093/eurpub/cki155.
21. Eser E, Lağarlı T, Baydur H, et al. [Pschometric properties of the Turkish version of the EUROHIS-Tr (WHOQOL-8-Tr) in a Turkish population]. Turk J Public Health. 2011;8(3):136-152. [Article in Turkish]
22. Koopman C, Pelletier KR, Murray JF, et al. Stanford Presenteeism Scale: Health status and employee productivity. J Occup Environ Med. 2002;44(1):14-20. doi: 10.1097/00043764-200201000-00004.
23. Baysal İA, Baysal G, Aksu G, Aksu N. [The Relation Between Presenteeism (The Problem of Not Attending Work) and Organızational Commitment: A Study on The Academic Staff of Adnan Menderes University]. Ejovoc (Electronic Journal of Vocational Colleges). 2014;4(3):134-152. [Article in Turkish]
24. Coşkun Ö, Akdur Recep. [Evaluation of Effective Factors in Absenteeism and Presenteeism at Two Workplaces]. In: Temizer A, Baytal Y, editors. New Trends in Socıal Scıences-V. Podgorica, Montenegro. 2018. pp. 415-448. [Article in Turkish]
25. Dalkılıç E. Stanford işte varolamama ölçeğinin Türkçe'de geçerlik ve güvenirliği: Hemşireler üzerinde bir araştırma. [Master's thesis]. İstanbul University; 2017. [Article in Turkish].
26. İnceoğlu F, Demir P, Aydoğdu H. Adaptation of Fear of Missing Out Scale (FoMOs) to Dentistry. Selcuk Dent J. 2021;8(2):530-537. doi: 10.15311/selcukdentj.946507.
27. Batmaz H, Ulusoy Y, İnceoğlu F. The Mediating Role Of Digital Game Addiction In The Correlation Between Cyber Victimization And Cyber Bullying. Int Soc Sci Stud J. 2020;6(73):5093-5108. doi: 10.26449/sssj.2726.
28. Gliem AJ, Gliem RR. Calculating, interpreting and reporting Cronbach's alpha reliability coefficient for Likert-type scales. 2003 Midwest Research to Practice Conference in Adult, Continuing, and Community Education. 2003:82-88.
29. Arbuckle JL. IBM SPSS AMOS 21 User's Guide. IBM Corp; 2012.
30. İnceoğlu F, Yoloğlu S, Kanık EA. Comparative analysis of models in confirmatory factor analysis: Exploring clinical applications and interpretation. Med Sci. 2023;12(2):562-568. doi: 10.5455/medscience.2022.12.278.
31. Gürbüz S. Sosyal Bilimlerde Aracı ve Düzenleyici Etki Analizleri (3rd ed.). Seçkin Yayıncılık:, Ankara. 2025. [Article in Turkish].
32. İnceoğlu F, Porgalı-Zayman E. Moderator effect of chronic disease on the relationship between marriage adjustment and satisfaction in married couples. Eur J Ther. 2023;29(3):459-468. doi: 10.58600/eurjther1722.
33. Chen Y, Guo K, Ji Q, Zhang D. Not all climate risks are alike: Heterogeneous responses of financial firms to natural disasters in China. Finance Res Lett. 2023;52:103538. doi: 10.1016/j.frl.2022.103538.
34. Türkiye Earthquakes Recovery and Reconstruction Assessment. Presidency of the Republic of Turkey, Presidency of Strategy and Budget. Available at: https://www.sbb.gov.tr/wp-content/uploads/2023/03/Turkiye-Recovery-and-Reconstruction-Assessment.pdf. Accessed September 01, 2024.
35. Taşkın Ö. [Health Economics in Sustainable Development and Analysis of Edirne Province in Terms of Health Sector]. Master's thesis. Trakya University; 2011. [Article in Turkish]
36. Arslaner E, Boylu Y. [Presenteeism in Work Life: An Evaluation in Hotel Industry]. İşletme Araştırmaları Derg. 2015;7(4):123-136. [Article in Turkish].
37. Canbaz S. [Investigation of the Presenteeism Levels of Business Employees in Terms of Demographıc Characteristics]. Sosyal Bilimler Metinleri. 2021;2:25-41. [Article in Turkish].
38. Stewart WF, Ricci JA, Chee E, Morganstein D, Lipton R. Lost productive time and cost due to common pain conditions in the US workforce. JAMA. 2003;290(18):2443-2454. doi: 10.1001/jama.290.18.2443.
39. Hemp P. Presenteeism: at work--but out of it. Harv Bus Rev. 2004;82(10):49-58, 155.
40. Yılmaz G. [The Effect of Presenteeism and Job Stress on Employee Performance: An Application Example]. Master's thesis. Trakya University; 2019. [Article in Turkish]
41. Karaçınar H. [The Effect of Employees' Dark Leadership and Favoritism Perceptions on Presenteeism: A Research on Textile Workers]. Journal of Economics Business and Political Researches. 2022;7(19):620-643. doi: 10.25204/iktisad.1148684. [Article in Turkish]
42. Çetinkaya B. [The Mediating Effect of Work Engagement

and Moderating Effect of Work Efficacy in the Relationship Between Presenteeism and Absenteeism: ISO 500 Research]. PhD thesis. Selçuk University; 2024. [Article in Turkish].
43. İncekara İ. [The role of presenteeism and mindfulness in the impact of job demands and job resources on job performance]. PhD thesis. Başkent University; 2024. [Article in Turkish].
44. Bulan S, Söyük S. [Evaluation of Theses on Presenteism]. Int J Discipl Econ Adm Sci Stud. 2022;8(46):721-730. doi: 10.29228/ideas.64501. [Article in Turkish].
45. Özmen G. [The relationship between presenteeism and organizational commitment: A study on textile workers]. Master's thesis. Eskişehir Osmangazi University; 2011. [Article in Turkish].
46. Howard JT, Howard KJ. The effect of perceived stress on absenteeism and presenteeism in public school teachers. J Workplace Behav Health. 2020;35(2):100-116. doi: 10.1080/15555240.2020.1724794.
47. Li Y, Zhang J, Wang S, Guo S. The Effect of Presenteeism on Productivity Loss in Nurses: The Mediation of Health and the Moderation of General Self-Efficacy. Front Psychol. 2019;10:1745. doi: 10.3389/fpsyg.2019.01745.
48. Masuda M, Ishimaru T, Hino A, et al; CORoNaWork Project. A Cross-Sectional Study of Psychosocial Factors and Sickness Presenteeism in Japanese Workers During the COVID-19 Pandemic. J Occup Environ Med. 2022 Jan 1;64(1):e1-7. doi: 10.1097/JOM.0000000000002415.
49. Magalhaes S, Barbosa J, Borges E. The relationship between presenteeism, quality of life and social support in higher education professionals: A cross-sectional path analysis. PLoS One. 2022;17(4):e0267514. doi: 10.1371/journal.pone.0267514.
50. Aronsson G, Hagberg J, Björklund C, et al. Health and motivation as mediators of the effects of job demands, job control, job support, and role conflicts at work and home on sickness presenteeism and absenteeism. Int Arch Occup Environ Health. 2021;94(3):409-418. doi: 10.1007/s00420-020-01591-w.

The European Research Journal 2025;11(5):957-965 Original Article

DOI: https://doi.org/10.18621/eurj.1707845 *Obstetrics and Gynecology*

Fixed- versus adjustable-loop femoral fixation in anterior cruciate ligament surgery: Is one method clinically superior?

Mücahid Osman Yücel, Raşit Emin Dalaslan, Sönmez Sağlam, Zekeriya Okan Karaduman, Mehmet Arıcan

Department of Orthopaedics and Traumatology, Düzce University, Faculty of Medicine, Düzce, Türkiye

ABSTRACT

Objectives: This study aimed to compare fixed-loop and adjustable-loop femoral fixation systems in anterior cruciate ligament (ACL) reconstruction to determine whether one provides superior short-term clinical outcomes.

Methods: This retrospective cohort study included 88 patients who underwent anatomic ACL reconstruction by the same surgical team at Düzce University Faculty of Medicine between January 01, 2018 and December 31, 2023. Patients were divided into two groups as those with a fixed-loop (n=39) and those with an adjustable-loop (n=49). A standard postoperative rehabilitation program was applied to all patients. Lysholm knee score, International Knee Documentation Committee (IKDC) score, Visual Analog Scale (VAS), and range of motion (ROM) were evaluated preoperatively and 6 months postoperatively as clinical evaluation parameters.

Results: Both groups demonstrated significant improvement in VAS, Lysholm, and IKDC scores during the postoperative period (P<0.001). Postoperative Lysholm score was 77.5±14.8 and IKDC score was 64.4±10.7 in the fixed-loop group, while these values were 78.4±4.1 and 66.7±10.3 in the adjustable-loop group, respectively. There was no significant difference in postoperative scores between the groups (P>0.05).

Conclusions: Fixed and adjustable-loop femoral fixation systems used in ACL reconstruction show similar performance in terms of short-term clinical outcomes. Both methods can be safely preferred; the final choice of method should be based on the surgeon's experience and patient-specific factors. Further studies evaluating long-term outcomes are needed.

Keywords: Anterior cruciate ligament, reconstruction, orthopedic fixation devices, treatment outcome, arthroscopy, range of motion

Anterior cruciate ligament (ACL) tears represent one of the most frequently encountered knee pathologies, particularly in physically active and younger populations, posing a significant threat to joint stability and functional performance [1]. ACL reconstruction (ACLR) is a surgical method that is widely used today, especially in athletes to return to their former activity level. In these surgeries, fixation

Received: May 29, 2025 **Accepted:** August 22, 2025 **Available Online:** August 25, 2025 **Published:** September 4, 2025

How to cite this article: Yücel MO, Dalaslan RE, Sağlam S, Karaduman ZO, Arıcan M. Fixed- versus adjustable-loop femoral fixation in anterior cruciate ligament surgery: Is one method clinically superior? Eur Res J. 2025;11(5):957-965. doi: 10.18621/eurj.1707845

Corresponding author: Mücahid Osman Yücel, MD., Assist. Prof., Phone: +90 380 542 13 90 ext. 6555, E-mail: mucahidosmanyucel@gmail.com

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

of the graft with adequate stability on the femoral and tibial sides is critical to ensure biological integration of the graft with bone during the healing process [2]. One of the most preferred fixation methods on the femoral side is cortical suspension systems [3, 4]. These systems are divided into two main groups: fixed-loop and adjustable-loop [5, 6].

Although fixed-loop systems have high load-bearing capacity and good biomechanical properties, they do not always provide complete placement of the graft in the tunnel, which may lead to movement of the graft in the tunnel and synovial fluid leakage, increasing the risk of tunnel enlargement [7, 8]. In contrast, adjustable-loop systems offer the advantage of accommodating various tunnel lengths and are designed to minimize micromotion by ensuring a tighter fit of the graft within the tunnel [9, 10]. However, some biomechanical studies have suggested an increased risk of loop loosening and graft slippage after use in these systems [11].

Clinical studies in recent years have focused on comparing fixed and adjustable-loop systems in terms of functional outcomes. In these studies, clinical evaluation criteria such as Lysholm, IKDC and VAS scores were generally used and no significant difference was found between the two systems in most studies [12]. However, some studies have reported differences, especially in terms of radiologic parameters. It has been shown that postoperative complications, especially bone tunnel widening, may vary according to the fixation technique [13, 14].

Although both fixed and adjustable fixation systems are widely used in ACLR, their comparative effectiveness regarding clinical outcomes remains insufficiently clarified. Therefore, our study aimed to evaluate the clinical outcome of patients treated with fixed-loop versus adjustable-loop devices and to assess how the choice of fixation method may influence functional recovery.

METHODS

In the retrospective cohort study, patients who underwent anatomic ACLR at Düzce University Faculty of Medicine between January 01, 2018 and December 31, 2023 was evaluated. Ethics committee approval was granted by Düzce University Medical Faculty Clinical Research Ethics Committee with decision number 2024/128 dated 21.04.2025. Owing to the retrospective nature of the study, the need for informed consent was waived. All surgical procedures were performed by the same surgical team using standard surgical technique, and the same postoperative rehabilitation protocol was applied to all patients.

In the study, patients were classified into two groups: fixed-loop fixation system (Loopfix Button Ti Femoral Fixation Devices) and adjustable-loop fixation system (Liftfix Button Ti Femoral Fixation Devices) (Fig. 1). Visual Analog Scale (VAS), International Knee Documentation Committee (IKDC) score, Lysholm knee score, and range of motion were evaluated preoperatively and 6 months postoperatively as clinical evaluation parameters.

Statistical Analysis

All analyses were conducted using IBM SPSS Statistics v20.0. The Shapiro-Wilk test assessed normality. Depending on distribution, Independent Samples t-test or Mann-Whitney U test was used for intergroup comparisons, and Paired Samples t-test or Wilcoxon signed-rank test for pre–post comparisons. Data were summarized as mean±standard deviation or median (minimum-maximum). Statistical significance was set at $P < 0.05$, and effect sizes were calculated using Cohen's d. A post-hoc power analysis was performed for the primary between-group comparisons (postoperative Lysholm and IKDC scores) using a two-sample t-test with a two-sided $\alpha=0.05$. The observed effect sizes were small (Lysholm d=0.09; IKDC d=0.22), corresponding to differences of less than 2-3 points between groups. With the current sample size (n=39 vs. n=49), the power to detect such small effects was limited (6.9% for Lysholm; 17.3% for IKDC). However, the study had adequate sensitivity to detect moderate or larger differences ($d \geq 0.61$, ≈ 6 points), suggesting that clinically meaningful moderate-to-large effects are unlikely to have been missed.

RESULTS

A total of 88 patients (n=39/49) were included in the study. Six of the patients were female. The study evaluated the effect of fixed-loop fixation systems and adjustable-loop fixation systems on clinical parameters

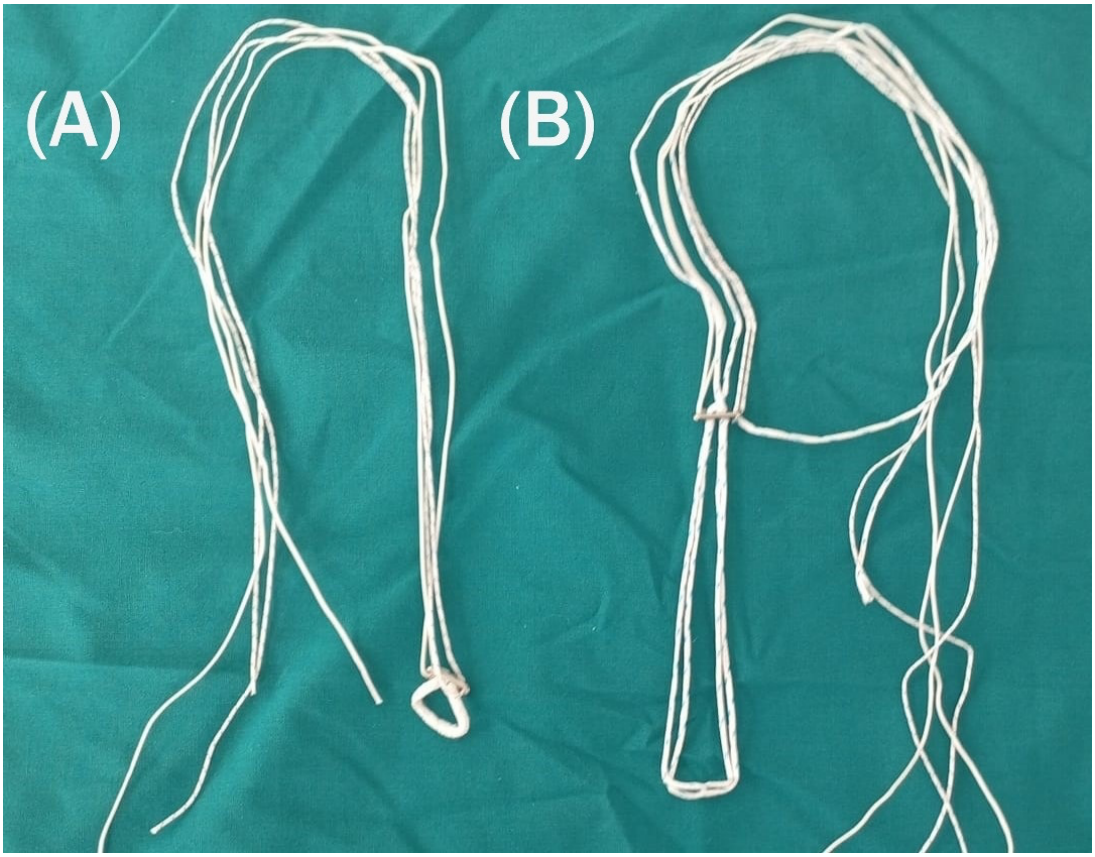

Fig. 1. (A) Fixed-loop (Loopfix Button Ti) and (B) adjustable-loop (Liftfix Button Ti) femoral fixation devices used in the study.

in patients undergoing ACLR.

As shown in Table 1, when preoperative and postoperative values were compared in all patients, statistically significant improvements were observed in VAS, Lysholm, and IKDC scores in the postoperative period (P<0.001; Wilcoxon, Paired Samples t-Test, respectively). A significant increase in range of motion (ROM) flexion values was also found (P<0.001), while the improvement in ROM extension was not statistically significant (P=0.131).

These findings suggest that the surgical intervention was clinically and functionally effective. In addition to the statistical significance, effect size analysis revealed a large effect for IKDC (d= –1.88) and a moderate-to-large effect for Lysholm (d= –0.76). A large effect size was also observed for the reduction

Table 1. Comparison of preoperative and postoperative scores within the entire study group

Parameter	Preoperative	Postoperative	P value	Effect size
VAS	1.8±1.7	0.5±0.9	**<0.001**[a]	-0.597[g]
Lysholm	62.7±18.1	78.0±10.3	**<0.001**[c]	-0,755[f]
IKDC	41.4±7.5	65.7±10.5	**<0.001**[c]	-1,883[f]
ROM (flexion)	112.1±19.4	120.8±10.0	**<0.001**[a]	-0.379[g]
ROM (extension)	1.5±1.9	1.1±1.3	<0.131[a]	-0.161[g]

Data are shown as mean±standard deviation. VAS=Visual Analog Scale, IKDC=International Knee Documentation Committee, ROM=Range of Motion

[a]Wilcoxon Signed-Rank Test, [c]Paired Samples t-Test, [f]Cohen's d, [g]Rank-biserial r

Table 2. Comparison of postoperative values between fixed-loop and adjustable-loop groups

Parameter	Fixed-Loop Postoperative	Adjustable-Loop Postoperative	P-value	Effect size
VAS	0.6±1.0	0.5±0.8	0.460[b]	-0.079[g]
Lysholm	77.5±14.8	78.4±4.1	0.736[d]	-0.077[f]
IKDC	64.4±10.7	66.7±10.3	0.321[d]	-0.215[f]
ROM (flexion)	118.7±11.0	122.5±8.8	0.146[b]	-0.155[g]
ROM (extension)	1.4±1.4	0.9±1.2	0.072[b]	-0.192[g]

Data are shown as mean±standard deviation. VAS=Visual Analog Scale, IKDC=International Knee Documentation Committee, ROM=Range of Motion

[b]Mann-Whitney U Test, [d]Independent Samples t-Test, [f]Cohen's d, [g]Rank-biserial r

in VAS scores (r= –0.60). ROM flexion showed a moderate effect (r= –0.38), while ROM extension exhibited only a small effect size (r= –0.16).

Table 2 compares the postoperative values between the fixed-loop and adjustable-loop groups. There was no statistically significant difference between the groups in postoperative VAS (P=0.460), Lysholm (P=0.736), and IKDC (P=0.321) scores. Similarly, no significant difference was observed in postoperative ROM flexion (P=0.146) and ROM extension (P=0.072) values (Kruskal-Wallis test).

Although none of the comparisons reached statistical significance, the effect size analysis revealed small effect sizes across all parameters. According to the effect size measures, Cohen's d values were small for Lysholm (d= –0.077) and IKDC (d= –0.215), while rank-biserial r values also indicated small effects for VAS (r= –0.079), ROM flexion (r= –0.155), and ROM extension (r= –0.192).

The findings show that both fixation methods resulted in significant improvements in clinical and functional outcomes compared to the preoperative period (Fig. 2). However, there was no statistically significant difference in postoperative scores between the

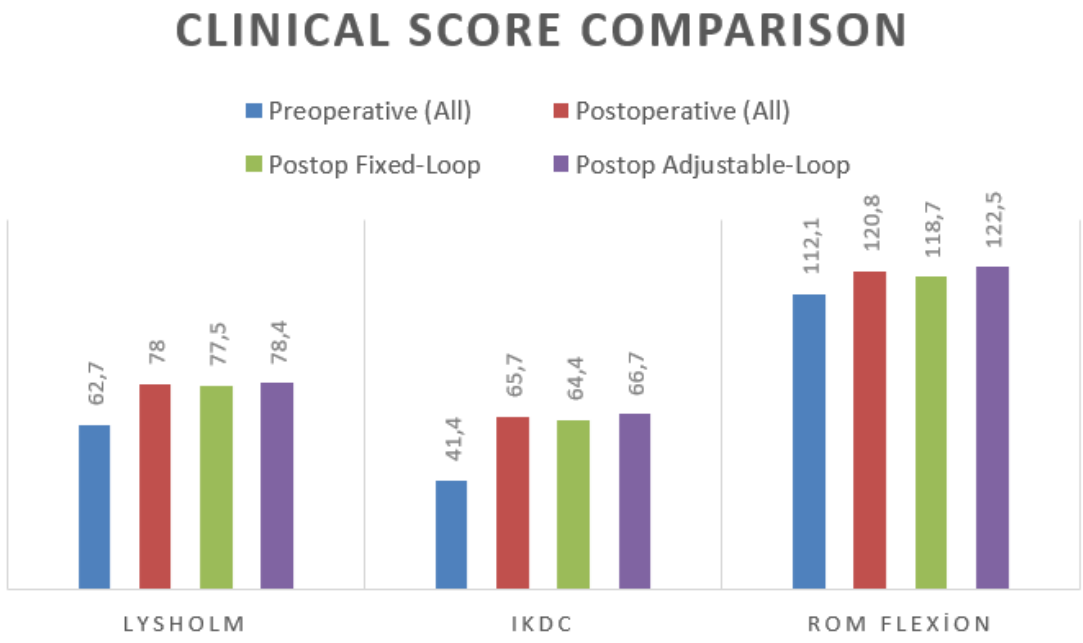

Fig 2. Comparison of preoperative and postoperative clinical scores between all patients, fixed-loop, and adjustable-loop groups. Lysholm Knee Score, International Knee Documentation Committee (IKDC) score, and range of motion (ROM) flexion values are shown. Postop=postoperative.

groups. These results suggest that both fixed and adjustable-loop systems are effective methods and preferences should be shaped according to surgical experience and patient characteristics.

DISCUSSION

The clinical outcomes of fixed-loop and adjustable-loop femoral fixation systems in ACLR were compared. The findings revealed that both systems provide similar success in terms of postoperative functional outcomes.

The differences between fixed and adjustable-loop systems are important not only clinically but also in terms of biomechanical and radiologic parameters. There are several studies showing the advantages of fixed-loop systems. Choi *et al*. [15] reported that ACLRs performed with loop systems allowed less tunnel widening and better graft integration. Similarly, some other studies in the literature reported that fixed-loop devices caused less graft loosening and maintained knee stability better [16, 17]. These findings suggest that fixed systems can provide mechanical stability more effectively.

On the other hand, some advantages of adjustable-loop systems have also been emphasized in the literature. Studies by Eichinger *et al*. [7] and Hyodo *et al*. [18] showed that adjustable systems can limit tunnel expansion and may positively affect graft healing by filling the femoral tunnel better. In contrast, Bachmaier *et al*. [19] reported that adjustable-loop systems tend to loosen over time, which may have negative effects on clinical performance. Smith *et al*. [20] stated that both systems provide similar biomechanical stability, but some adjustable systems may require re-tensioning and loss of force depending on the design. In Kocazeybek *et al*.'s [21] prospective study, it was observed that the application of the adjustable system with different surgical techniques did not make a significant difference on knee stability and objective measurements. In addition, Kyriakidis *et al*'s [22] study in a group of young patients reported that the double adjustable system achieved successful radiologic and functional results. This is thought to be related to the inability of the graft to completely fill the femoral tunnel in fixed-loop systems and to allow micro-movement of the graft within the tunnel. In particular, it is suggested that the mechanisms of intra-graft mobility defined as "bungee effect" and "windshield wiper effect" contribute to this expansion.

In conclusion, fixed and adjustable cortical suspension systems each have their own advantages. Studies comparing these two systems in the literature are still current and research is ongoing to strengthen these comparisons based on different patient groups, surgical techniques and follow-up periods. In our study, we found that these two systems were not superior to each other in terms of clinical outcomes.

Previous studies have reported a wide range of Lysholm scores following ACL reconstruction, with variability depending on the fixation method. For example, Onggo *et al*. [4] found fixed-loop systems yielded scores between 82.3 and 93.9, while adjustable-loop systems ranged from 85.7 to 94.7, and Mohamed *et al*. [23] reported almost identical results (90.15 vs. 90.62). Similarly, Choi *et al*. [15] and Sheth *et al*. [17] found no significant difference between the two systems, whereas Hyodo *et al*. [18] observed substantial improvement with adjustable fixation. IKDC score comparisons also revealed similar trends; studies by Sheth *et al*. [17], Kocazeybek *et al*. [21], and Eichinger *et al*. [7] consistently showed no significant difference, while Ahn *et al*. [24] and Kaya *et al*. [25] reported variable baseline values but comparable postoperative improvements. A summary of these results is provided in Table 3, illustrating the consistent finding of equivalent short-term functional outcomes between fixed-loop and adjustable-loop femoral fixation methods across multiple studies.

In this study, there was no significant difference between fixed and adjustable systems in terms of Lysholm and IKDC scores. This finding suggests that both techniques are effective in providing functional stability of the knee and increasing patient satisfaction. In this direction, it can be said that fixed and adjustable systems give similar results in terms of clinical efficacy, and the preferred method should be shaped according to the surgeon's experience and patient-specific factors.

From a clinical perspective, the absence of significant differences between the two femoral fixation methods suggests that surgeons can choose either system without compromising short-term patient outcomes. For the patient, this means that functional recovery and pain

Table 3. Summary of previous studies comparing fixed-loop and adjustable-loop femoral fixation in anterior cruciate ligament reconstruction

Author, Year	Fixation type(s)	Lysholm	IKDC	Main Finding
Onggo *et al.*, 2019 [4]	Fixed vs. Adjustable (systematic review)	Fixed: 82.3-93.9; Adjustable: 85.7-94.7	–	No significant difference
Mohamed *et al.*, 2020 [23]	Fixed vs. Adjustable	Fixed: 90.15; Adjustable: 90.62	–	No significant difference
Choi *et al.*, 2017 [15]	Fixed vs. Adjustable	Fixed: 58.3±16.6; Adjustable: 58.1±16.2	–	No significant difference
Hyodo *et al.*, 2023 [18]	Adjustable	Preop: 82.9±10.5; Postop: 99.6±1.4	–	Significant improvement
Sheth *et al.*, 2019 [17]	Fixed vs. Adjustable	Fixed: 94.23; Adjustable: 94.32	Fixed: 92.03; Adjustable: 92.16	No significant difference
Kocazeybek *et al.*, 2023 [21]	Adjustable (3 techniques)	95-98	88.8-94.9	No significant difference
Eichinger *et al.*, 2023 [7]	Fixed vs. Adjustable	Fixed: 73-90; Adjustable: 77-91	Fixed: 60±8→92±6; Adjustable: 65±17→88±17	Both improved significantly
Ahn *et al.*, 2018 [24]	Adjustable	–	78.6±17.7	No significant difference
Kaya *et al.*, 2025 [25]	Fixed vs. Adjustable	Fixed: 83.19; Adjustable: 81.35	Fixed: 66.69; Adjustable: 63.44	No significant difference
Asif *et al.*, 2021 [12]	Fixed vs. Adjustable	Adjustable: 91.0 ± 3.6; Fixed: 91.4 ± 3.5	Adjustable: 91.9 ± 3.6; Fixed: 91.5 ± 3.6	No significant difference

Data are shown as mean±standard deviation or median (minimum-maximum). IKDC=International Knee Documentation Committee, Preop=preoperative, Postop=postoperative

relief are likely to be similar regardless of the chosen device, allowing other factors such as graft availability, cost, surgical familiarity, and intraoperative considerations to guide decision-making. This flexibility may be particularly valuable in settings where device availability or budget constraints limit options.

In our study, the improvement in ROM extension was small and did not reach statistical significance. This finding may be related to multiple factors rather than the fixation method itself. Variations in patient compliance with postoperative rehabilitation protocols could influence the degree of extension recovery. Minor measurement inaccuracies, especially when assessing small extension deficits, may also contribute to variability.

Limitations

One of the most important limitations of our study is that the evaluations were limited to subjective clinical scores and objective knee stability measurements (e.g., arthrometer) were not performed. Furthermore, control of variables such as graft quality, biological factors such as individual bone density, and individual compliance with rehabilitation programs is limited. Future research should focus on long-term follow-up to determine whether the absence of significant differences between fixed-loop and adjustable-loop femoral fixation persists over time. Prospective multicenter trials with larger sample sizes and inclusion of objective stability measurements (e.g., arthrometer testing, radiologic assessment of tunnel widening) would provide more robust evidence to guide clinical decision-making.

CONCLUSION

In conclusion, fixed and adjustable femoral fixation systems used in ACLR demonstrate comparable short-term clinical outcomes. Both techniques are considered safe and effective, and the final choice should be guided by the surgeon's experience and the specific needs of the patient.

Ethics Approval and Consent to Participate

This study was approved by the Düzce Uıniversity Clinical Research Ethics Committee (Decision No: 2024/128; date: 21.04.2025). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Informed consent was waived because of the retrospective nature of the study and the analysis used anonymous clinical data.

Data Availability

All data relevant to this study are included in the article. Additional data can be obtained from the corresponding author upon reasonable request.

Authors' Contribution

Study Conception: MOY, RED; Study Design: MOY, SS; Supervision: MOY, ZOK; Funding: N/A; Materials: MA, ZOK; Data Collection and/or Processing: RED, SS; Statistical Analysis and/or Data Interpretation: MA, SS; Literature Review: ZOK, RED; Manuscript Preparation: MOY, MA; and Critical Review: ZOK, SS.

Conflict of Interest

The authors disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The authors disclosed that they did not receive any grant during the conduction or writing of this study.

Acknowledgments

The authors have no acknowledgments to declare.

Generative Artificial Intelligence Statement

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

Editor's Note

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

REFERENCES

1. Motififard M, Akbari Aghdam H, Ravanbod H, et al. Demographic and Injury Characteristics as Potential Risk Factors for Anterior Cruciate Ligament Injuries: A Multicentric Cross-Sectional Study. J Clin Med. 2024;13(17):5063. doi: 10.3390/jcm13175063.
2. Yao S, Fu BS, Yung PS. Graft healing after anterior cruciate ligament reconstruction (ACLR). Asia Pac J Sports Med Arthrosc Rehabil Technol. 2021;25:8-15. doi: 10.1016/j.asmart.2021.03.003.
3. Rodríguez-Merchán EC. Anterior Cruciate Ligament Reconstruction: Is Biological Augmentation Beneficial? Int J Mol Sci. 2021;22(22):12566. doi: 10.3390/ijms222212566.
4. Onggo JR, Nambiar M, Pai V. Fixed- Versus Adjustable-Loop Devices for Femoral Fixation in Anterior Cruciate Ligament Reconstruction: A Systematic Review. Arthroscopy. 2019;35(8):2484-2498. doi: 10.1016/j.arthro.2019.02.029.
5. Ranjan R, Gaba S, Goel L, et al. In vivo comparison of a fixed loop (EndoButton CL) with an adjustable loop (TightRope RT) device for femoral fixation of the graft in ACL reconstruction: A prospective randomized study and a literature review. J Orthop Surg (Hong Kong). 2018;26(3):2309499018799787. doi: 10.1177/2309499018799787.
6. Özdemir G, Turan S, Aslan H, Bingöl O, Deveci A, Kılıç E. Anterior cruciate ligament reconstruction with ToggleLoc with ZipLoop system versus transfix system: A cost-effectiveness analysis. Ulus Travma Acil Cerrahi Derg. 2021;27(2):231-237. doi: 10.14744/tjtes.2020.22487.
7. Eichinger M, Ploner M, Degenhart G, et al. Tunnel widening after ACL reconstruction with different fixation techniques: aperture fixation with biodegradable interference screws versus all-inside technique with suspensory cortical buttons. 5-year data from a prospective randomized trial. Arch Orthop Trauma Surg. 2023;143(11):6707-6718. doi: 10.1007/s00402-023-05001-x.
8. Lin Y, Zhang L, Shen S, et al. No Difference in Bone Tunnel Enlargement and Clinical Outcome between Cortical Suspension and Hybrid Femoral Fixation in Hamstring Anterior Cruciate Ligament Reconstruction. Orthop Surg. 2024;16(4):902-911. doi: 10.1111/os.14024.
9. Lee TJ, Jang KM, Kim TJ, Lee SM, Bae JH. Adjustable-Loop Cortical Suspensory Fixation Results in Greater Tibial Tunnel Widening Compared to Interference Screw Fixation in Primary Anterior Cruciate Ligament Reconstruction. Medicina (Kaunas). 2022;58(9):1193. doi: 10.3390/medicina58091193.
10. Singh S, Shaunak S, Shaw SCK, Anderson JL, Mandalia V. Adjustable Loop Femoral Cortical Suspension Devices for Anterior Cruciate Ligament Reconstruction: A Systematic Review. Indian J Orthop. 2020;54(4):426-443. doi: 10.1007/s43465-019-00022-4.
11. Houck DA, Kraeutler MJ, McCarty EC, Bravman JT. Fixed- Versus Adjustable-Loop Femoral Cortical Suspension Devices for Anterior Cruciate Ligament Reconstruction: A Systematic Review and Meta-analysis of Biomechanical Studies. Orthop J Sports Med. 2018;6(10):2325967118801762. doi: 10.1177/2325967118801762.
12. Asif N, Khan MJ, Haris KP, Waliullah S, Sharma A, Firoz D. A prospective randomized study of arthroscopic ACL reconstruction with adjustable- versus fixed-loop device for femoral side fixation. Knee Surg Relat Res. 2021;33(1):42. doi: 10.1186/s43019-021-00124-0.
13. McDermott E, DeFoor MT, Blaber OK, Aman ZS, DePhillipo NN, Dekker TJ. Biomechanical comparison of anterior cruciate ligament reconstruction fixation methods and implications on clinical outcomes. Ann Jt. 2023;8:15. doi: 10.21037/aoj-22-52.
14. Robinson JR, Haddad FS. ACL graft failure: surgical technique may affect outcomes. Bone Joint J. 2021;103-B(9):1439-1441. doi: 10.1302/0301-620X.103B9.BJJ-2021-0919.
15. Choi NH, Yang BS, Victoroff BN. Clinical and Radiological Outcomes After Hamstring Anterior Cruciate Ligament Reconstructions: Comparison Between Fixed-Loop and Adjustable-Loop Cortical Suspension Devices. Am J Sports Med. 2017;45(4):826-831. doi: 10.1177/0363546516674183.
16. Chandru V, MS S, JS ST, Nair RR. Comparison of Fixed- and Variable-Loop Button Fixation in Arthroscopic Anterior Cruciate Ligament Reconstruction. Cureus. 2022;14(4):e24218. doi: 10.7759/cureus.24218.
17. Sheth H, Salunke AA, Barve R, Nirkhe R. Arthroscopic ACL reconstruction using fixed suspensory device versus adjustable suspensory device for femoral side graft fixation: What are the outcomes? J Clin Orthop Trauma. 2019;10(1):138-142. doi: 10.1016/j.jcot.2017.09.004.
18. Hyodo K, Kanamori A, Watanabe R, et al. Does loop length change after anterior cruciate ligament reconstruction with adjustable loop cortical suspension device?: Observation of the hamstring graft completely filling the femoral tunnel. J Exp Orthop. 2023;10(1):67. doi: 10.1186/s40634-023-00629-5.
19. Bachmaier S, Monaco E, Smith PA, Frank RM, Matzkin EG, Wijdicks CA. Biomechanical Comparison of 3 Adjustable-Loop Suspensory Devices for All-Inside ACL Reconstruction: A Time-Zero Full-Construct Model. Orthop J Sports Med. 2023;11(9):23259671231201461. doi: 10.1177/23259671231201461.
20. Smith PA, Piepenbrink M, Smith SK, Bachmaier S, Bedi A, Wijdicks CA. Adjustable- Versus Fixed-Loop Devices for Femoral Fixation in ACL Reconstruction: An In Vitro Full-Construct Biomechanical Study of Surgical Technique-Based Tibial Fixation and Graft Preparation. Orthop J Sports Med. 2018;6(4):2325967118768743. doi: 10.1177/2325967118768743.
21. Kocazeybek E, Meric E, Ersin M, et al. Clinical outcomes of three different techniques using adjustable-loop fixation in arthroscopic single-bundle anterior cruciate ligament reconstruction: A prospective randomized clinical trial. Knee. 2023;43:208-216. doi: 10.1016/j.knee.2023.06.015.
22. Kyriakidis T, Tzaveas A, Melas I, Petras K, Iosifidou AM, Iosifidis M. Anterior Cruciate Ligament Reconstruction Utilizing Double Adjustable-Loop Suspensory Fixation Devices Provides Good Clinical Outcomes in Patients under the Age of 40 Years at Two-Year Follow-Up. J Clin Med. 2024;13(18):5436. doi: 10.3390/jcm13185436.
23. Mohamed RSAA, El-Din El-Shafie MH, El-Sheikh MA. Clinical Outcome of Fixed Versus Adjustable Loop Cortical Suspension Devices in Arthroscopic Anterior Cruciate Ligament Reconstruction. Ortop Traumatol Rehabil. 2020;22(3):181-185. doi: 10.5604/01.3001.0014.3234.

24. Ahn JH, Ko TS, Lee YS, Jeong HJ, Park JK. Magnetic Resonance Imaging and Clinical Results of Outside-in Anterior Cruciate Ligament Reconstruction: A Comparison of Fixed- and Adjustable-Length Loop Cortical Fixation. Clin Orthop Surg. 2018;10(2):157-166. doi: 10.4055/cios.2018.10.2.157.

25. Kaya S, Guven N, Unal YC, et al. Comparison of long-term clinical and radiologic outcomes of AperFix and fixed loop device fixation in anterior cruciate ligament reconstruction: A retrospective study. Medicine (Baltimore). 2025;104(3):e41199. doi: 10.1097/MD.0000000000041199.

The European Research Journal 2025;11(5):966-976 Original Article

DOI: https://doi.org/10.18621/eurj.1731334 *Orthopedics and Traumatology*

Posterior calcaneal spur length and angle are predictors of pain and functional limitation in insertional Achilles tendinopathy

Mustafa Dinç[1], Ömer Cevdet Soydemir[1], Recep Karasu[1], Bilal Aykaç[1], Hünkar Çağdaş Bayrak[2]

[1]Department of Orthopedics and Traumatology, Bursa City Hospital, Bursa, Türkiye; [2]Department of Orthopedics and Traumatology, Çekirge State Hospital, Bursa, Türkiye

ABSTRACT

Objectives: This study aimed to quantitatively evaluate the clinical impact of posterior calcaneal spur (PoCS) morphology, specifically spur length and inclination angle, in patients with insertional Achilles tendinopathy (IAT).

Methods: This retrospective study analyzed 200 patients with symptomatic IAT who underwent standardized weight-bearing lateral ankle radiographs. Spur length and inclination angle were measured, and patients were stratified into nine subgroups based on three length categories (<5 mm, 5-10 mm, >10 mm) and three angle categories (<10°, 10-20°, >20°). Clinical outcomes were assessed using the Visual Analog Scale (VAS), American Orthopaedic Foot and Ankle Society (AOFAS) score, and Victorian Institute of Sport Assessment-Achilles (VISA-A) score. Kruskal-Wallis tests and multivariate linear regression analyses were used to evaluate associations between spur morphology and outcomes.

Results: Both longer spurs (>10 mm) and steeper inclination angles (>20°) were significantly associated with worse clinical scores including higher VAS scores and lower AOFAS and VISA-A scores (P<0.001). Patients with spur lengths >10 mm and angles >20° had a mean VAS score of 7.22±0.65, VISA-A score of 49.72±2.54, and AOFAS score of 60.00±4.24, indicating greater pain and functional limitation. In contrast, patients with spur lengths <5 mm and angles <10° had lower VAS scores (5.18±0.82) and higher VISA-A (63.43±3.92) and AOFAS (72.57±4.33) scores, reflecting lower pain intensity and higher functional capacity (P < 0.001 for all). Regression analysis confirmed that spur length and angle were independent predictors of clinical outcome (P<0.001), while age, sex, and BMI were not statistically significant contributors (P>0.05).

Conclusions: Spur morphology - specifically length and angle - has a measurable impact on symptom severity in IAT. Radiographic evaluation of PoCS morphology should be integrated into clinical decision-making for more tailored management.

Keywords: Posterior calcaneal spur, insertional Achilles tendinopathy, spur angle, spur length

Posterior heel pain is a common musculoskeletal complaint, affecting a significant portion of the population, particularly middle-aged and older adults [1]. It accounts for up to 15% of all foot and ankle complaints in clinical practice [2, 3]. Among the various etiologies, insertional Achilles tendinopathy

Received: July 1, 2025 **Accepted:** August 2, 2025 **Available Online:** August 8, 2025 **Published:** September 4, 2025

How to cite this article: Dinç M, Soydemir ÖC, Karasu R, Aykaç B, Bayrak HÇ. Posterior calcaneal spur length and angle are predictors of pain and functional limitation in insertional Achilles tendinopathy. Eur Res J. 2025;11(5):966-976. doi: 10.18621/eurj.1731334

Corresponding author: Mustafa Dinç, MD., Phone: +90 224 975 00 00, E-mail: drindian@hotmail.com

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

(IAT) and posterior calcaneal spur (PoCS) formation are frequently implicated [3]. Studies estimate that PoCS is present in 65%-80% of patients with IAT but also occurs in 25%-35% of asymptomatic individuals [3-5]. This raises the critical question of whether spur morphology influences clinical symptoms or if these bony outgrowths are simply incidental findings.

PoCS forms at the Achilles tendon insertion, often due to chronic traction, calcification, and degenerative changes within the tendon [5, 6]. In contrast, Haglund's deformity is a bony prominence at the posterosuperior calcaneus that leads to impingement of the Achilles tendon and retrocalcaneal bursitis [7, 8]. While Haglund's deformity is primarily a mechanical compression issue, PoCS is more closely associated with insertional Achilles tendinopathy and enthesopathy [7, 8]. These differences necessitate distinct treatment approaches, yet both conditions may coexist and contribute to posterior heel pain.

Despite the high prevalence of PoCS in IAT, there is a lack of studies examining how the size and orientation of the spur affect clinical outcomes. Previous research on plantar calcaneal spurs (PCS) has demonstrated that spur length and slope correlate with pain severity and functional impairment, yet no study has systematically evaluated these factors for posterior calcaneal spurs [9-11].

This study aims to investigate the relationship between posterior calcaneal spur morphology (length and slope) and clinical outcomes in patients with insertional Achilles tendinopathy. To our knowledge, this is the first study to classify PoCS based on its morphological characteristics and assess its impact on pain, function, and Achilles tendon pathology. By establishing objective measurement criteria for PoCS, we aim to improve diagnostic accuracy, treatment decision-making, and patient outcomes in posterior heel pain syndromes.

METHODS

Study Design

The study was designed as a retrospective cohort investigation and performed at Bursa City Hospital, Department of Orthopedics and Traumatology. It adhered to the ethical principles outlined in the Declaration of Helsinki and received approval from the institutional ethics committee (approval number: 2025-5/3; date: 05.03.2025).

Patient Selection

This study included adult patients aged 18 years or older who presented to the Orthopedics and Traumatology outpatient clinic at Bursa City Hospital between January 2021 and December 2023 with posterior heel pain. Additional demographic and clinical variables - such as age, sex, body mass index (BMI), symptom duration, and physical activity level - were documented and included in the statistical models to control for potential confounding factors. All outcome measures were recorded during the same clinical visit as radiographic assessment to maintain temporal consistency.

Eligible patients were required to have a clinical diagnosis of IAT, supported by symptoms of chronic pain localized to the Achilles tendon insertion, tenderness upon palpation, and functional limitation during dorsiflexion or tiptoe walking [4, 5]. Radiographic confirmation of PoCS was mandatory, based on lateral weight-bearing ankle radiographs. Patients were included only if both clinical and radiographic criteria were met, and high-quality imaging was available for morphological analysis.

Patients were excluded if they exhibited radiographic signs of Haglund's deformity, defined as a prominent posterosuperior calcaneal exostosis, in order to isolate the effect of PoCS on IAT. Additional exclusion criteria included the presence of pre-insertional or non-insertional Achilles tendinopathy, a history of Achilles tendon rupture or previous surgery in the region, retrocalcaneal bursitis, or systemic inflammatory disorders such as rheumatoid arthritis or ankylosing spondylitis. Patients with diabetes mellitus, gout, neurological disorders, or peripheral neuropathies were also excluded due to their potential impact on tendon pathology. Furthermore, individuals with a history of corticosteroid injections or other invasive interventions targeting the Achilles tendon within the preceding six months were not included. After applying these criteria, a total of 200 patients were selected for the final cohort, ensuring that the study population represented isolated cases of insertional Achilles tendinopathy associated with posterior calcaneal spur formation, free from confounding mechanical or systemic influences.

Radiographic Evaluation and Spur Morphology Classification

All patients underwent standard lateral weight-bearing ankle radiographs to assess bony morphology at the Achilles tendon insertion. The radiographs were independently reviewed by two orthopedic surgeons who were blinded to clinical data and had at least five years of experience in musculoskeletal imaging interpretation.

PoCS were defined as osseous protrusions emanating from the posterior aspect of the calcaneus at the insertion site of the Achilles tendon. Special attention was given to distinguish PoCS from Haglund's deformity, which is characterized by a posterosuperior calcaneal prominence located superior and lateral to the Achilles insertion [6-8, 12].

Two morphological features of PoCS were evaluated: spur length and spur angle. Spur length was measured as the linear distance from the posterior cortical base of the calcaneus to the distal tip of the posterior calcaneal spur. For subgroup analysis, spur length was stratified into three categories: short (<5 mm), intermediate (5-10 mm), and long (>10 mm) (Fig. 1).

PoCS angle was defined as the angle formed between a line drawn from the tip of the spur to its base at the posterior calcaneal cortex and a second line tangent to the posterior surface of the calcaneus, representing the longitudinal axis of the calcaneal body. This angle was further classified into three groups based on inclination severity: mild (<20°), moderate (20°-30°), and severe (>30°) (Fig. 2). These thresholds were adapted from previous studies on PCS morphology, which used similar groupings to evaluate spur-related mechanical effects [9, 11]. All radiographic measurements were performed using calibrated digital software (Fonet DICOM Viewer, Version 4.1; Fonet Information Technologies, Ankara, Turkey), and inter-observer reliability was assessed to ensure consistency, with high agreement achieved through consensus in cases of discrepancy.

Outcome Measures

Clinical evaluation focused on both pain severity and functional impairment associated with insertional Achilles tendinopathy. To ensure a comprehensive assessment, three validated instruments were employed: the Visual Analog Scale (VAS) for pain intensity [13], the American Orthopaedic Foot and Ankle Society (AOFAS) Hindfoot Score for functional performance [14], and the Victorian Institute of Sport Assessment-Achilles (VISA-A) Questionnaire, which is specific to Achilles tendon-related disorders [15].

Pain intensity was measured using the VAS, a 10-cm horizontal line where patients marked their average pain experienced over the previous week. The score ranged from 0 (no pain) to 10 (worst imaginable pain), allowing for a simple yet reliable quantification of symptom severity.

The AOFAS Hindfoot Score was used to assess foot and ankle function, incorporating components of pain (40 points), function (50 points), and alignment (10 points) for a maximum of 100 points. This composite score captures both subjective patient feedback

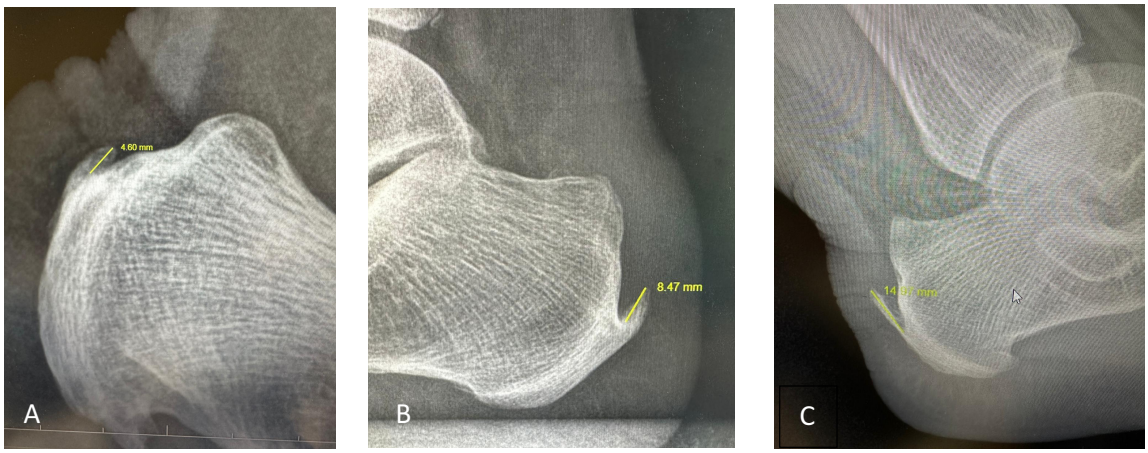

Fig. 1. Representative lateral radiographs illustrating posterior calcaneal spurs of varying lengths. A) Short spur (<5 mm), B) Intermediate spur (5-10 mm), C) Long spur (>10 mm). Spur length was measured from the calcaneal cortex to the tip of the bony outgrowth using standardized radiographic tools under weight-bearing conditions.

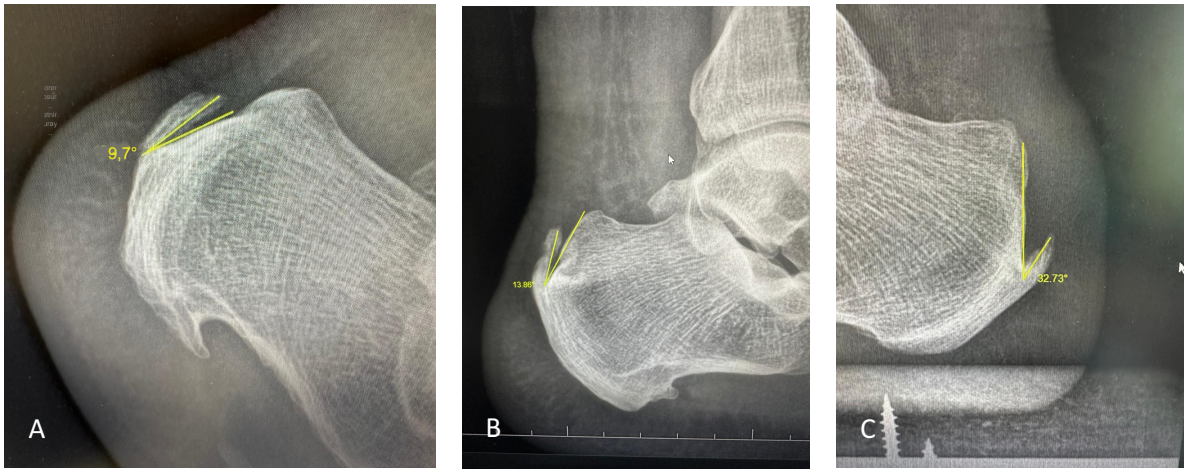

Fig. 2. Representative lateral radiographs demonstrating posterior calcaneal spur angles categorized by severity. A) Mild angle (<10°), B) Moderate angle (10-20°), C) Steep angle (>20°). Spur angle was measured as the angle between a line drawn from the spur tip to its base and a line tangent to the posterior surface of the calcaneus. All measurements were performed under standardized weight-bearing radiographic conditions.

and objective clinical findings. All clinical examinations contributing to the AOFAS were performed by two independent orthopedic specialists blinded to the radiographic spur morphology.

The VISA-A questionnaire was utilized to specifically evaluate symptoms and functional limitations due to Achilles tendinopathy. This validated tool consists of eight items assessing domains such as pain, function during daily living, and physical activity. The total VISA-A score ranges from 0 to 100, with higher scores indicating better Achilles tendon function and lower symptom burden. The VISA-A is especially sensitive in detecting clinical changes in both athletic and non-athletic populations suffering from insertional or midportion Achilles tendinopathy.

Statistical Analysis

All statistical analyses were conducted using IBM SPSS Statistics for Windows, Version 26.0 (IBM Corp., Armonk, NY, USA), with the level of statistical significance set at $P<0.05$. Descriptive statistics were expressed as means and standard deviations (SD) for continuous variables and as frequencies and percentages for categorical variables. The normality of continuous variables, including VAS, AOFAS, and VISA-A scores, was evaluated using the Shapiro–Wilk test and Q–Q plots. Since the normality assumption was not fully satisfied for several outcome variables, non-parametric tests were employed for group comparisons where appropriate. To determine independent predictors of pain intensity and functional limitation, multiple linear regression analyses were performed using VAS and AOFAS scores as dependent variables. Independent variables entered into the model included age, sex, BMI, symptom duration, spur length, spur angle. Multicollinearity was checked using the variance inflation factor (VIF), with all values remaining below 2.0, indicating acceptable levels. Additionally, to evaluate the interaction between spur size and angle, patients were stratified into nine morphological subgroups based on combinations of spur length (<5 mm, 5-10 mm, >10 mm) and spur angle (<20°, 20°-30°, >30°). The Kruskal–Wallis test was used to compare VAS, AOFAS, and VISA-A scores across these subgroups due to the non-parametric nature of the data. Post hoc pairwise comparisons were adjusted using Bonferroni correction to control for Type I error. All subgroup results were graphically displayed as boxplots for visual inspection of score distributions and potential outliers.

RESULTS

Table 1 displays the demographic and clinical characteristics of patients stratified into nine subgroups based on PoCS length (<5 mm, 5-10 mm, >10 mm) and spur angle (<10°, 10°-20°, >20°). For each subgroup, av-

Table 1. Combined distribution of posterior calcaneal spur (PoCS) length and angle categories with clinical and demographic parameters

Groups	Variables	<10°	10-20°	>20°
<5 mm	**Age**	43.93±9.02	45.58±9.83	42.33±3.66
	BMI	26.15±2.74	24.80±2.72	27.20±2.78
	Duration	13.89±1.93	13.50±1.41	13.89±1.64
	Spur length	3.96±0.47	4.36±0.31	3.88±0.68
	Spur angle	8.26±1.24	17.34±1.47	30.12±1.46
	VISA-A	63.43±3.92	61.00±3.68	62.11±3.34
	AOFAS	72.57±4.33	73.42±3.11	72.44±3.60
	VAS	5.18±0.82	5.58±0.50	5.89±0.32
5-10 mm	**Age**	43.08±8.65	42.67±6.81	44.80±9.95
	BMI	25.31±2.90	25.90±2.54	26.09±1.98
	Duration	14.15±1.43	14.67±1.27	14.30±1.53
	Spur length	8.80±0.71	8.88±0.73	8.78±0.85
	Spur angle	7.80±1.17	16.58±1.18	30.63±1.57
	VISA-A	63.23±2.27	60.67±1.83	60.70±2.98
	AOFAS	71.85±3.32	73.13±3.97	70.35±3.50
	VAS	5.54±0.51	5.67±0.56	5.75±0.55
>10 mm	**Age**	46.77±5.24	44.45±11.25	47.00±10.85
	BMI	25.25±2.27	26.54±2.65	24.95±2.32
	Duration	14.00±1.63	14.70±1.89	17.33±1.61
	Spur length	13.69±0.74	13.74±0.72	13.80±0.70
	Spur angle	8.09±1.34	17.46±1.40	31.13±1.83
	VISA-A	62.14±3.03	60.20±4.53	49.72±2.54
	AOFAS	69.86±5.48	66.90±7.06	60.00±4.24
	VAS	6.14±0.64	6.35±0.67	7.22±0.65

Data are shown as mean±standard deviation. Spur length was stratified into three categories: short (<5 mm), intermediate (5-10 mm), and long (>10 mm). Spur angle was classified into mild (<10°), moderate (10-20°), and severe (>20°) inclination. Clinical outcomes assessed include the Visual Analog Scale (VAS), Victorian Institute of Sports Assessment-Achilles questionnaire (VISA-A), and American Orthopaedic Foot & Ankle Society (AOFAS) hindfoot scores, BMI=Body Mass Index.

erage values are provided for age, BMI, symptom duration, spur length, and spur angle. Additionally, clinical outcome scores including VISA-A, AOFAS, and VAS are reported.

Multiple linear regression analyses were conducted to evaluate the relationship between PoCS morphology and clinical outcomes, including VAS, VISA-A, and AOFAS scores. For the VAS score, spur length (P<0.001) and spur angle (P<0.001) were found to be statistically significant positive predictors of pain severity. In contrast, patient age (P=0.112), BMI (P=0.525), and sex (P=0.963) had no significant influence (Table 2).

In the VISA-A score regression, spur length (P<0.001) and spur angle (P<0.001) were again significant, demonstrating a negative correlation with functional status. Neither age (P=0.560), BMI (P=0.453), nor sex (P=0.114) had a statistically significant effect (Table 3).

Similarly, for AOFAS scores, longer spur length (P<0.001) and larger spur angles (P=0.001) were associated with significantly lower functional outcomes.

Table 2. Multivariate linear regression analysis identifying predictors of VAS score

Variables	t value	P value	95% Confidence Interval
Age	-1.596	0.112	-0.018-0.002
BMI	-0.636	0.525	-0.045-0.023
Spur angle	5.846	0.0	0.046-0.092
Spur angle	5.746	0.0	0.019-0.038
Sex	-0.046	0.963	-0.183-0.175

Values are based on multivariate linear regression analysis. VAS=Visual Analog Scale, BMI=Body Mass Index. Significant predictors of pain severity include spur length and spur angle (P<0.001), while age, BMI, and sex were not statistically significant.

Age (P=0.503), BMI (P=0.133), and sex (P=0.917) did not exhibit a meaningful relationship with AOFAS scores (Table 4).

Subgroup analyses by Spur Morphology, patients were stratified into nine subgroups based on spur length (<5 mm, 5-10 mm, >10 mm) and spur angle (<10°, 10°–20°, >20°) to further explore morphological effects on outcomes. Group-wise comparisons using the Kruskal-Wallis test revealed significant differences in VAS, VISA-A, and AOFAS scores across the subgroups. Patients with spur lengths >10 mm and angles >20 exhibited the most severe symptoms, with VAS scores reaching a median of 7.5-8.0, VISA-A scores decreasing to around 50, and AOFAS scores falling to approximately 60. In contrast, those with spurs <5 mm and angles <10 showed milder symptoms, with VAS scores near 5.0, VISA-A scores of 65-70, and AOFAS scores of 75-80 (Fig. 3).

Table 3. Multivariate linear regression analysis for predictors of VISA-A score

Variables	t value	P value	95% Confidence Interval
Age	-0.584	0.56	-0.079-0.043
BMI	-0.752	0.453	-0.282-0.127
Spur angle	-4.499	0.0	-0.463-0.181
Spur angle	-6.295	0.0	-0.249-0.130
Sex	1.587	0.114	-0.212-1.956

VISA-A=Victorian Institute of Sports Assessment-Achilles questionnaire, BMI=Body Mass Index. Increased spur length and angle were independently associated with lower VISA-A scores (P<0.001), indicating worse symptom severity and reduced tendon function. Age, BMI, and sex were not statistically significant predictors

Table 4. Multivariate linear regression analysis for predictors of AOFAS score

Variables	t value	P value	95% Confidence Interval
Age	0.67	0.503	-0.051-0.103
BMI	-1.509	0.133	-0.457-0.061
Spur length	-6.717	0.0	-0.787-0.429
Spur angle	-3.322	**0.001**	-0.202-0.051
Sex	0.104	0.917	-1.299-1.444

AOFAS=American Orthopaedic Foot & Ankle Society Hindfoot Score, BMI=Body Mass Index. Both increased spur length and angle were independently associated with significantly lower AOFAS scores (P<0.01), reflecting reduced ankle-hindfoot function. Age, BMI, and sex were not significant predictors.

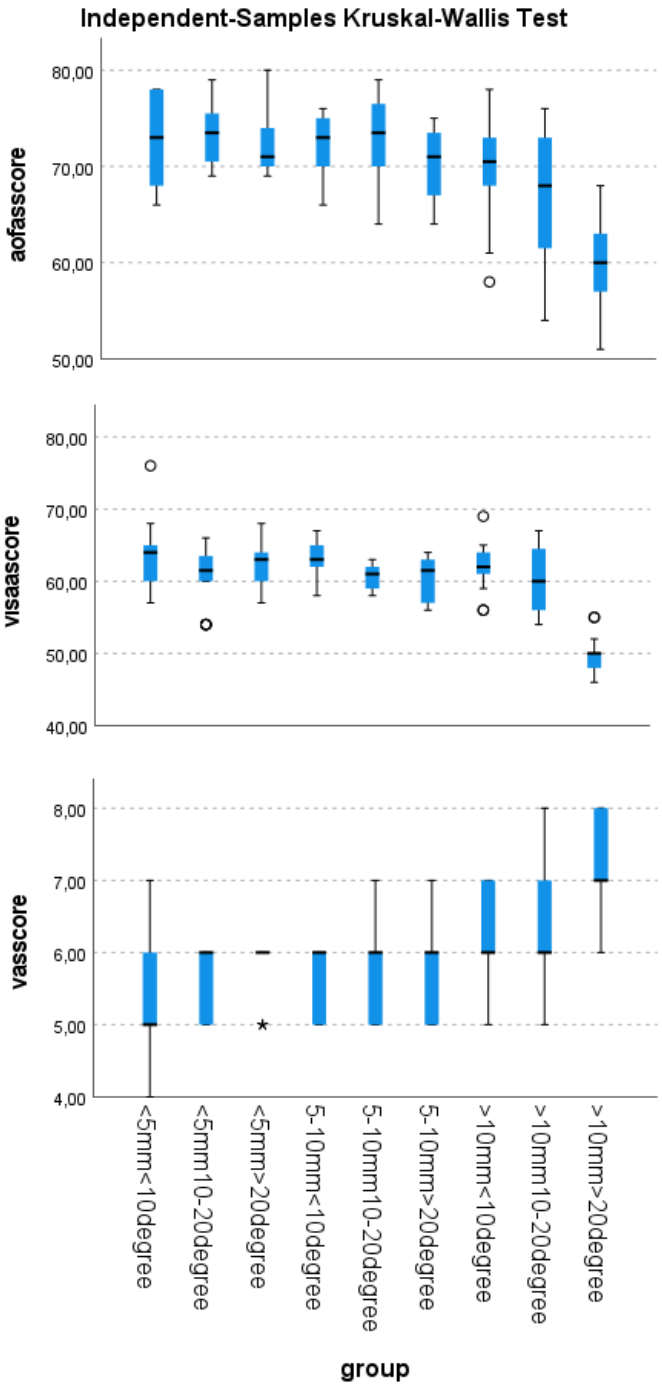

Fig. 3. Comparison of clinical outcome scores across posterior calcaneal spur morphology subgroups. Box plots depicting AOFAS, VISA-A, and VAS scores stratified by posterior calcaneal spur length (<5 mm, 5-10 mm, >10 mm) and angle (<10°, 10-20°, >20°). The Kruskal-Wallis test indicated statistically significant differences among groups for all three clinical outcomes. Patients with spur lengths >10 mm and angles >20° had the lowest AOFAS and VISA-A scores and the highest VAS scores, reflecting worse function and greater pain severity. Medians are shown by horizontal lines, boxes represent interquartile ranges, whiskers indicate full ranges, and circles/asterisks denote outliers.

DISCUSSION

Our analysis revealed that both spur length and spur angle were significant independent predictors of pain and function. Specifically, multiple linear regression results demonstrated a positive correlation between VAS scores and both spur length and angle, indicating that patients with longer and more vertically inclined spurs experience greater pain severity. In contrast, demographic factors such as age, BMI, or sex did not show a significant association with pain levels.

Similarly, the VISA-A and AOFAS functional scores exhibited a significant inverse relationship with spur length and angle. As the size and inclination of the PoCS increased, functional scores decreased, suggesting that more prominent spurs compromise tendon insertion mechanics and contribute to functional limitations. These trends remained consistent across both scoring systems. Again, demographic parameters had no statistically significant influence on outcomes.

The subgroup analysis, stratifying patients into nine morphological categories based on combined spur length and angle, further confirmed these associations. Patients with extreme morphology (>10 mm length and >20° angle) consistently showed the highest VAS scores and the lowest VISA-A and AOFAS scores, highlighting the additive impact of both variables. Conversely, those with small (<5 mm) and shallow (<10°) spurs had better clinical outcomes.

These findings are reinforced by existing biomechanical insights. These relationships can be explained through established biomechanical mechanisms. Chimenti *et al*. [5, 16] emphasized the complex loading patterns at the Achilles enthesis, where the superficial fibers are subjected to tensile stress while the deep fibers experience compressive forces from the calcaneus. A steeply inclined PoCS (>30°) may amplify these compressive forces, creating a sharper mechanical interface at the deep tendon insertion. This intensified contact area can promote localized stress accumulation, microtears, inflammation, and collagen disorganization. In contrast, milder angles may permit more uniform load transmission, potentially reducing focal mechanical irritation. Thus, our observation that steeper angles are linked to more severe symptoms supports the hypothesis that specific angular morphologies critically disrupt enthesis mechanics and exacerbate degenerative changes.

Although not directly focused on posterior calcaneal spurs or insertional Achilles tendinopathy, two recent studies have evaluated plantar calcaneal spur morphology and its clinical relevance. Kaya *et al*. [17] demonstrated that the mere presence of a calcaneal spur is not necessarily a primary determinant of heel pain, implying that spur morphology - such as length and angle -may be more clinically meaningful. Similarly, Tuncer *et al*. [18] found that larger plantar calcaneal spurs and greater spur–calcaneus angles were significantly correlated with higher pain levels and worse functional scores in patients with plantar heel pain treated with ESWT. Although these studies primarily addressed plantar spurs and plantar fasciitis, the mechanistic insights they provide regarding the role of spur morphology in symptom severity are consistent with and indirectly support our findings in posterior calcaneal spurs.

In addition to angular effects, spur size appears to play a mechanical role. Larger PoCS (>10 mm) may project deeper into the tendon's insertional zone, increasing mechanical impingement and friction during ankle dorsiflexion - especially under load-bearing conditions. Chimenti *et al*. [5, 16] reported that symptomatic heels exhibit significantly longer spurs compared to asymptomatic ones, reinforcing the link between increased spur length and pathological mechanical interactions. This chronic mechanical interference likely contributes to enthesis overload and a failed healing response, leading to degenerative features such as neovascularization, fatty infiltration, and tendon thickening - hallmarks of chronic tendinopathy. As noted by Krishna Sayana and Maffulli [6] and Boberg and Anania [12], such enthesophytes may initially arise as adaptive responses to repetitive tensile stress but may evolve into pathological bony projections when combined with altered joint mechanics.

The anatomical concept of a calcaneal "step," introduced by Fiamengo *et al* [19]., may offer further insight. This bony protrusion, situated at the tendon insertion, may function as a fulcrum that modifies the tendon's line of pull and increases mechanical tension during motion. Although its morphology was not specifically analyzed in terms of length or angle, its frequent appearance in symptomatic individuals highlights the clinical relevance of posterior calcaneal morphology. Additionally, Grambart *et al* [7]. and Caudell *et al* [8]. emphasized the importance of distinguishing PoCS from Haglund's deformity, noting that PoCS is more directly implicated in intratendinous degeneration and tendon entrapment, whereas Haglund's deformity primarily affects the retrocalcaneal bursa.

Although previous studies have examined the presence of posterior calcaneal spurs and their relationship with Achilles tendinopathy, no prior research has directly investigated the correlation between PoCS length and angle in relation to clinical outcomes. Kang *et al*. [4] reported that calcification width ranged from 1.0 to 10.4 mm and length from 3.5 to 16.2 mm in insertional Achilles tendinopathy, suggesting that a broader and longer calcific deposit may increase mechanical irritation. However, they did not directly analyze the correlation between width, length, and clinical symptoms, making it unclear whether larger calcifications necessarily lead to worse functional impairment. Furthermore, they found that Haglund's deformity was not indicative of insertional Achilles tendinopathy and was equally present in asymptomatic patients. Their study also confirmed that a majority of insertional Achilles tendinopathy patients had calcification at the tendon insertion, reinforcing the association between calcification and IAT but not providing a detailed assessment of how size impacts symptom severity. Lu *et al*. [20] analyzed posterior calcaneal step spurs and Achilles tendon calcifications in Haglund syndrome, showing that 56.8% of symptomatic patients had a posterior calcaneal step spur, compared to only 5% in the control group. Their study reinforces the importance of osseous variations in posterior heel pain, but it did not assess the impact of PoCS length or angle on clinical outcomes, which is the focus of our study. Fiamengo *et al*. [19] introduced the concept of the calcaneal step, an anatomical variation at the Achilles tendon insertion observed in symptomatic patients with posterior heel pain. Their findings indicated that a calcaneal step was present in 75% of symptomatic cases, suggesting that this anatomical feature may contribute to Achilles-related pathology. However, their study did not specifically assess the correlation between posterior calcaneal spur length, width, or angle and clinical outcomes, nor did it provide a biomechanical explanation for how the calcaneal step influences pain and function. Our study expands upon these findings by demonstrating the direct impact of PoCS on clinical outcomes, emphasizing the importance of spur morphology beyond simple

anatomical variations. Nakajima [21] conducted a radiographic comparison of symptomatic and asymptomatic heels in patients with insertional Achilles tendinopathy and examined both calcification length and width. Their study primarily focused on differentiating calcifications seen in symptomatic patients from those in asymptomatic individuals, rather than directly correlating spur dimensions with pain and function. Notably, they did not find a clear association between PoCS width and Achilles tendinopathy severity, nor did they analyze spur angles. Their research was more focused on distinguishing Haglund's deformity measurements, rather than assessing the detailed impact of PoCS morphology on clinical outcomes. Johansson *et al*. [22] analyzed posterior calcaneal spur size and surgical outcomes, finding that while spur size was a key anatomical feature, it did not correlate with surgical success rates. However, their study did not examine the direct correlation between spur morphology and clinical symptoms, which our study addresses. Grambart *et al*. [7] distinguished Achilles insertional calcific tendinosis (AICT) from Haglund's deformity, emphasizing that AICT involves intratendinous calcifications at the Achilles insertion, whereas Haglund's deformity is a separate mechanical compression issue. However, they did not analyze PoCS morphology or its impact on clinical outcomes, a gap that our study fills. Caudell [8] highlighted the importance of radiographic imaging in distinguishing PoCS from Haglund's deformity, emphasizing that IAT is frequently misdiagnosed. Their study confirmed that PoCS often presents alongside Achilles tendon thickening and intratendinous calcifications, yet their analysis did not explore how PoCS morphology influences clinical symptoms. In contrast, our study provides a quantitative analysis of PoCS length and angle, directly correlating these parameters with patient-reported outcomes.

Taken together, these mechanistic insights align with our findings and support the notion that not only the presence but the detailed morphology of posterior calcaneal spurs plays a pivotal role in the pathophysiology and clinical severity of IAT. Recognizing the influence of both spur length and inclination angle may be essential in clinical evaluation, guiding prognosis, and informing surgical decision-making, particularly in patients who do not respond to conservative treatment.

Strengths and Limitations

This study has several strengths that enhance the robustness and clinical relevance of its findings. First, to our knowledge, it is the first to quantitatively assess the impact of PoCS length and angle on clinical outcomes using validated patient-reported outcome measures (VAS, VISA-A, AOFAS). The incorporation of detailed radiographic measurements allows for objective stratification of spur morphology, moving beyond simple presence or absence toward a more nuanced understanding of structural contribution to symptoms. Second, the subgroup analysis offers granular insight by exploring the additive effects of combined spur length and angle variations on pain and function.

However, several limitations should be acknowledged. The retrospective design may introduce selection bias and limit causal inference. While radiographic measurements were standardized, inter-observer variability in angle and length assessment is a potential source of measurement error. Furthermore, although patient-reported outcome measures provide valuable information, they are inherently subjective and may be influenced by external psychosocial factors. Advanced imaging modalities such as MRI or ultrasound were not routinely employed, which could have offered complementary insight into soft tissue pathology such as tendon degeneration, neovascularization, or bursal inflammation. Additionally, routine assessments of lower extremity alignment (genu varus/valgus) and additional foot deformities (e.g., pes planus, hallux valgus) were not investigated; thus, their potential role as confounding factors influencing clinical outcomes remains unknown and warrants investigation in future prospective studies. Lastly, the study population was drawn from a single tertiary center, which may limit generalizability to broader patient populations.

The present findings underscore the importance of incorporating detailed PoCS morphological assessment into the diagnostic and therapeutic algorithm for IAT. Clinicians often focus on tendon pathology or generalized bony abnormalities; however, this study shows that specific morphologic traits - particularly spur length >10 mm and angle >30° are significantly associated with greater symptom burden. Radiographic identification of such features may help in stratifying patients likely to respond poorly to conservative treatment and potentially benefit from early sur-

gical intervention. Additionally, recognition of moderate-to-severe angular deformities could guide surgical planning, including decisions about whether to resect the spur or modify the tendon's insertion angle to restore favorable biomechanics.

Future prospective studies with larger, multicenter cohorts are needed to validate these findings and explore their prognostic utility. Incorporating advanced imaging techniques such as high-resolution ultrasound or MRI could help correlate PoCS morphology with tendon quality, neovascularization, or peritendinous inflammation. Additionally, biomechanical modeling or cadaveric studies may elucidate how different spur angles and sizes influence Achilles tendon loading, enthesis strain, and local tissue remodeling. Another avenue worth investigating is the impact of targeted surgical correction of spur morphology, particularly angle modification or step resection, on long-term pain relief and functional recovery. Such research would further clarify whether these radiographic parameters are merely markers of severity or modifiable contributors to disease progression.

CONCLUSION

In conclusion, this study demonstrates that both the length and inclination angle of posterior calcaneal spurs are independent predictors of clinical severity in insertional Achilles tendinopathy. Patients with longer and steeper spurs experience higher pain levels and reduced function, likely due to increased mechanical stress and tendon impingement at the enthesis. These results suggest that detailed radiographic assessment of PoCS morphology should be an integral part of clinical evaluation and may help guide individualized treatment strategies. By shifting the focus from presence to morphologic characterization, clinicians can more accurately predict outcomes and tailor interventions for patients with IAT.

Ethics Approval and Consent to Participate

This study was approved by the Bursa City Hospital Scientific Research Ethics Committee (Decision No: 2025-5/3; date: 05.03.2025). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Written informed consent was obtained from all individual participants included in the study.

Data Availability

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

Authors' Contribution

Study Conception: MD, ÖCS; Study Design: MD, RK; Supervision: MD, BA; Funding: MD, RK; Materials: MD, RK; Data Collection and/or Processing: MD, BA; Statistical Analysis and/or Data Interpretation: MD, HÇB; Literature Review: MD, BA; Manuscript Preparation: MD, BA and Critical Review: MD, RK.

Conflict of Interest

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

Financing

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

Acknowledgments

The authors have no acknowledgments to declare.

Generative Artificial Intelligence Statement

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

Editor's Note

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

REFERENCES

1. Chatterton BD, Muller S, Roddy E. Epidemiology of posterior

heel pain in the general population: cross-sectional findings from the clinical assessment study of the foot. Arthritis Care Res (Hoboken). 2015;67(7):996-1003. doi: 10.1002/acr.22546.
2. Lawrence DA, Rolen MF, Morshed KA, Moukaddam H. MRI of heel pain. AJR Am J Roentgenol. 2013;200(4):845-855. doi: 10.2214/ajr.12.8824.
3. Barg A, Ludwig T. Surgical Strategies for the Treatment of Insertional Achilles Tendinopathy. Foot Ankle Clin. 2019;24(3):533-559. doi: 10.1016/j.fcl.2019.04.005.
4. Kang S, Thordarson DB, Charlton TP. Insertional Achilles tendinitis and Haglund's deformity. Foot Ankle Int. 2012;33(6):487-491. doi: 10.3113/fai.2012.0487.
5. Chimenti RL, Cychosz CC, Hall MM, Phisitkul P. Current Concepts Review Update: Insertional Achilles Tendinopathy. Foot Ankle Int. 2017;38(10):1160-1169. doi: 10.1177/1071100717723127.
6. Krishna Sayana M, Maffulli N. Insertional Achilles tendinopathy. Foot Ankle Clin. 2005;10(2):309-320. doi: 10.1016/j.fcl.2005.01.010.
7. Grambart ST, Lechner J, Wentz J. Differentiating Achilles Insertional Calcific Tendinosis and Haglund's Deformity. Clin Podiatr Med Surg. 2021;38(2):165-181. doi: 10.1016/j.cpm.2020.12.003.
8. Caudell GM. Insertional Achilles Tendinopathy. Clin Podiatr Med Surg. 2017;34(2):195-205. doi: 10.1016/j.cpm.2016.10.007.
9. Şahin R, Sabri Balik M. Does the slope and length of the plantar calcaneal spur affect the clinic? Acta Orthop Belg. 2023;89(1):146-151. doi: 10.52628/89.1.10881.
10. Bayrak HÇ, Topaktaş S. [Comparative Outcomes of Percutaneous Plantar Fascia Release, Calcaneal Drilling, and Calcaneal Spur Removal versus ESWT in the Treatment of Plantar Fasciitis]. Osmangazi J Med. 2025;47(3):457-466. doi: 10.20515/otd1634034. [Article in Turkish]
11. Kuyucu E, Koçyiğit F, Erdil M. The association of calcaneal spur length and clinical and functional parameters in plantar fasciitis. Int J Surg. 2015;21:28-31. doi: 10.1016/j.ijsu.2015.06.078.
12. Boberg JS, Anania MC. Retrocalcaneal exostosis. Anatomy and a new surgical approach. J Am Podiatr Med Assoc. 2002;92(10):537-542. doi: 10.7547/87507315-92-10-537.
13. Delgado DA, Lambert BS, Boutris N, et al. Validation of Digital Visual Analog Scale Pain Scoring With a Traditional Paper-based Visual Analog Scale in Adults. J Am Acad Orthop Surg Glob Res Rev. 2018;2(3):e088. Epub 20180323. doi: 10.5435/JAAOSGlobal-D-17-00088.
14. SooHoo NF, Shuler M, Fleming LL. Evaluation of the validity of the AOFAS Clinical Rating Systems by correlation to the SF-36. Foot Ankle Int. 2003;24(1):50-55. doi: 10.1177/107110070302400108.
15. Robinson JM, Cook JL, Purdam C, et al. The VISA-A questionnaire: a valid and reliable index of the clinical severity of Achilles tendinopathy. Br J Sports Med. 2001;35(5):335-341. doi: 10.1136/bjsm.35.5.335.
16. Chimenti RL, Flemister AS, Tome J, et al. Altered tendon characteristics and mechanical properties associated with insertional achilles tendinopathy. J Orthop Sports Phys Ther. 2014;44(9):680-689. doi: 10.2519/jospt.2014.5369.
17. Kaya Ö, Demirtaş Y. [Is the presence radiologically calcaneal spur a factor for heel pain?] MKÜ Tıp Derg. 2022;13(47):281-285. doi: 10.17944/mkutfd.996223. [Article in Turkish]
18. Tuncer T, Karakeci ES. Effectiveness of extracorporeal shock wave therapy according to plantar calcaneal spur types. Ann Med Res. 2023;30(5):549-53. doi: 10.5455/annalsmedres.2022.12.372.
19. Fiamengo SA, Warren RF, Marshall JL, Vigorita VT, Hersh A. Posterior heel pain associated with a calcaneal step and Achilles tendon calcification. Clin Orthop Relat Res. 1982;(167):203-211.
20. Lu CC, Cheng YM, Fu YC, Tien YC, Chen SK, Huang PJ. Angle analysis of Haglund syndrome and its relationship with osseous variations and Achilles tendon calcification. Foot Ankle Int. 2007;28(2):181-185. doi: 10.3113/fai.2007.0181.
21. Nakajima K. Insertional Achilles tendinopathy: A radiographic cross-sectional comparison between symptomatic and asymptomatic heel of 71 patients. Eur J Radiol Open. 2024;12:100568. doi: 10.1016/j.ejro.2024.100568.
22. Johansson KJ, Sarimo JJ, Lempainen LL, Laitala-Leinonen T, Orava SY. Calcific spurs at the insertion of the Achilles tendon: a clinical and histological study. Muscles Ligaments Tendons J. 2013;2(4):273-277.

The European Research Journal 2025;11(5):977-988 Original Article

DOI: https://doi.org/10.18621/eurj.1742729 *Cardiology*

Impact of simplified transcatheter aortic valve implantation approach on procedural and clinical outcomes

Murat Can Güney

Department of Cardiology, Atılım University, Faculty of Medicine, Medicana International Ankara Hospital, Ankara, Türkiye

ABSTRACT

Objectives: Increased operator experience and development of techniques have reduced complication rates, and the new trend is the simplified approach (SA) in transcatheter aortic valve implantation (TAVI). The aim of this study is to compare the safety, efficacy, and outcomes of the SA with the standard non-simplified approach (NSA).

Methods: We retrospectively included 517 consecutive symptomatic severe aortic stenosis (AS) patients undergoing TAVI. The procedure is performed under general anesthesia accompanied by TEE and with predilatation in the NSA group. Whereas sedation and local anesthesia, removal of the routine use of transesophageal echocardiography (TEE), skipping the predilatation step in appropriate patients is adopted in SA group.

Results: Among 517 patients, 144 underwent TAVI with SA and 363 with NSA. The NSA group was treated with the most Sapien XT valve (69.4% vs. 92.8; P<0.001). There were no significant differences in post-procedural complications between the groups as defined by the Valve Academic Research Consortium (VARC)-2 criteria. Although there was a trend toward lower mortality at 30-day favoring SA group, this finding did not differ significantly between the groups (0% vs. 2.9%, respectively for SA and NSA groups, P=0.058). However, total cumulative mortality at the end of the follow-up period was found to be significantly reduced in the SA group (7.6% vs. 35.7; P<0.001). The multivariate logistic regression analysis revealed that predilatation, general anesthesia, TEE guidance, and simplified approach were independent predictors of total mortality.

Conclusions: Our study showed that simplified TAVI procedure was safe and was no related to adverse events. Compared to the NSA group, SA-TAVI had statistically significant lower total mortality rates.

Keywords: Transcatheter aortic valve implantation, minimal invasive, aortic stenosis, clinical outcomes

After its first introduction in 2002, the transcatheter aortic valve implantation (TAVI) procedure was quickly implemented and accepted throughout the world [1]. TAVI is the preferred way of treatment for patients with moderate to high-risk symptomatic severe aortic stenosis (AS), mainly where a transfemoral approach is possible [2, 4]. Increased operator experience, meticulous pre-procedural multimodality planning, development of vascular closure techniques have reduced complication rates, and the new trend is the simplified approach in TAVI [5].

Received: July 15, 2025 **Accepted:** August 20, 2025 **Available Online:** .August 26, 2025 **Published:** September 4, 2025

How to cite this article: Güney MC. Impact of simplified transcatheter aortic valve implantation approach on procedural and clinical outcomes. Eur Res J. 2025;11(5):977-988. doi: 10.18621/eurj.1742729

Corresponding author: Murat Can Güney, MD., Assist. Prof., Phone: +90 312 586 80 00, E-mail: murat_guney@msn.com

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

With the expansion of TAVI indications and the increase in the elderly population, operators shifted to simplified approaches without sacrificing safety and efficacy. Although simplified TAVI approaches have shown promising results in terms of safety and efficiency, current evidence is largely based on small-scale studies, and there remains a lack of sufficiently powered studies with large patient populations in the literature. In the simplified TAVI procedure, the transition from general anesthesia to sedation and local anesthesia, removal of the routine use of transesophageal echocardiography (TEE), skipping the predilatation step in appropriate patients and preference of the transfemoral access with percutaneous closure devices are the most critical steps. Although the first procedures were performed under general anesthesia and TEE, this is associated with the need for more positive inotropic agents, hemodynamic instability, a higher likelihood of pulmonary infections, challenges in extubation among patients with chronic lung disease, and extended duration of hospitalization [5]. Predilatation was considered as a crucial step in the development of the TAVI procedure. However, although predilatation has advantages such as reducing the radial forces and providing optimal device expansion, it should be remembered that it may cause complications such as the need for a permanent pacemaker, conduction disturbances, stroke, severe aortic regurgitation, and even annular rupture [6]. Furthermore, rapid pacing during predilatation leads to severe short-term hemodynamic instability at the procedure, which may lead to catastrophic outcomes in patients with depressed systolic functions [7]. Finally, incompatible with the simplified approach conventional technique may prolong the procedure time and increase the amount of contrast agent use.

Our study aimed to compare the safety, efficacy, and outcomes of the simplified approach (SA) with the standard non-simplified approach (NSA).

METHODS

Patient Population and Pre-procedural Planning

We retrospectively included 517 consecutive severe symptomatic AS patients undergoing TAVI using Edwards SAPIEN XT (Edwards Lifesciences, Irvine, CA, USA), SAPIEN S3 transcatheter heart valve (S3-THV) or The Lotus™ valve system (Boston Scientific, MA, USA) between July 2011 and June 2024. The heart team, consisting of one clinical cardiologist, one interventional cardiologist, one cardiovascular surgeon, discussed all patients and decided on treatment options. Baseline characteristics, laboratory and echocardiographic findings, procedural details, and outcome measures were retrospectively obtained. For all patients, aortic annulus size and vascular access site evaluated by multislice computerized tomographic angiography (MSCT) and echocardiography. Follow-up assessments were conducted post-procedure, at 30 days, and at 12 months. Final survival data were collected in June 2024. Written informed consent was obtained from all patients prior to the procedure, and the study was approved by the institutional ethics committee.

Procedure

The decision of the access site was made by individualized assessment in each patient, and the majority of patients underwent transfemoral access, while a smaller proportion were treated via the transaxillary route. Patients who underwent trans-axillary, transsubclavian access and valve-in-valve TAVI were excluded from the study. TAVI performed under general anesthesia accompanied by TEE and with predilation were included in the NSA group. None of the patients in the SA group had predilatation, and the procedure was performed with sedation, local anesthesia, and without TEE guidance. The vast majority of procedures were performed under deep sedation and local anesthesia. Predilatation decision before valve implantation was made by an experienced operator considering various factors. Routine predilatation involved the use of 20 mm, 23 mm, and 25 mm valvuloplasty balloons for the 23 mm, 26 mm, and 29 mm Edwards valves, and 22 mm and 24 mm balloons for the 25 mm and 27 mm Lotus valves, respectively. Predilatation balloon size in patients with Lotus valve implantation was decided not to exceed the minor axis length of the aortic annulus. In patients who underwent predilatation, a suitably sized Edwards or Lotus valvuloplasty balloon was advanced and inflated during rapid ventricular pacing. The balloon-expandable Edwards valve was then placed under rapid pacing, and fluoroscopic guidance, whereas the mechanically expand-

Table 1. Baseline characteristics and laboratory parameters

Parameters	All (n=517)	SA group (n=144)	NSA group (n=373)	P value
Age (years)	77.6±7.8	77.0±8.0	77.8±7.8	0.285
Female, n (%)	288 (55.7)	75 (52.1)	213 (57.1)	0.303
BMI (kg/m^2)	27.6±6.1	28.5±6.2	27.5±6.0	0.455
NYHA, n (%)				
2	138 (26.7)	44 (30.6)	94 (25.2)	
3	291 (56.3)	84 (58.3)	207 (55.35)	0.146
4	76 (14.7)	14 (9.7)	62 (16.6)	
Pulmonary edema	12 (2.3)	2 (1.4)	10 (2.7)	
DM, n (%)	154 (29.8)	50 (34.7)	104 (27.9)	0.127
HT, n (%)	426 (82.4)	122 (84.7)	304 (81.5)	0.389
HL, n (%)	258 (49.9)	79 (54.9)	179 (48.0)	0.161
CABG, n (%)	117 (22.6)	34 (23.6)	83 (22.3)	0.741
AF, n (%)	124 (24.0)	32 (22.2)	92 (24.7)	0.560
Stroke, n (%)	29 (5.6)	7 (4.9)	22 (5.9)	0.806
Previous MVR, n (%)	17 (3.2)	5 (3.4)	12 (3.2)	0.831
Moderate to severe COPD, n (%)	218 (42.1)	64 (44.4)	154 (41.2)	**0.010**
Renal replacement therapy, n (%)	12 (2.3)	7 (4.9)	5 (1.3)	**0.017**
STS score (%)	6.0±3.4	5.3±2.6	6.0±3.4	0.421
EuroSCORE II (%)	8.9±5.8	8.4±4.4	8.9±5.8	0.751
LogisticEUROSCORE (%)	23.6±9.3	22.6±10.8	25.4±7.8	0.788
CAD, n (%)				
Normal	169 (32.7)	42 (29.2)	127 (34.1)	0.223
Non-obstructive	220 (42.6)	50 (48.6)	70 (40.2)	
Obstructive	198 (24.8)	32 (22.2)	96 (25.7)	
Laboratory parameters				
Serum glucose (mg/dL)	127.2±54.6	132.4±61.9	125.2±51.4	0.179
Total cholesterol (mg/dL)	169.1±44.0	168.1±42.8	169.5±44.5	0.752
Triglyceride (mg/dL)	121.4±64.0	131.5±69.5	117.5±61.4	**0.027**
LDL cholesterol (mg/dL)	100.5±35.8	98.9±36.7	101.2±35.4	0.524
HDL cholesterol (mg/dL)	45.0±13.6	43.3±11.1	45.6±14.5	0.087
Creatinine (mg/dL)	1.0±0.53	1.0±0.36	1.1±0.58	0.547
Hemoglobin (mg/dL)	11.6±1.9	11.5±2.1	11.7±1.8	0.311
Platelet count (×10^3/L)	240.5±83.3	248.1±81.7	237.6±83.8	0.198
Troponin (pg/mL)	84.6±113.5	82.7±122.6	85.1±111.3	0.896
CK-MB (ng/mL)	4.4±11.0	3.5±3.6	4.6±12.3	0.493
Baseline echocardiographic parameters				
LVEF (%)	51.9±13.2	52.5±13.4	51.5±14.3	0.498
Bicuspid n (%)	69 (13.6)	14 (10.1)	55 (14.9)	0.156
Aortic velocity (cm/s)	4.4±0.61	4.2±0.58	4.5±0.61	**<0.001**
Aortic max gradient (mmHg)	82.3±23.1	74.8±20.0	85.2±23.6	**<0.001**
Aortic mean gradient (mm Hg)	50.6±15.1	45.5±12.4	52.5±15.6	**<0.001**
AVA (cm²)	0.67±0.16	0.69±0.16	0.66±0.16	**0.036**
Aortic Annulus (cm)	2.15±0.2	2.13±0.2	2.15±0.2	0.343
sPAP (mmHg)	44.1±16.7	44.4±17.7	44.0±16.3	0.784
Aortic regurgitation (moderate to severe), n (%)	21 (4.1)	6 (4.3)	15 (4.1)	0.058
Mitral regurgitation (moderate to severe), n (%)	63 (12.3)	24 (17.0)	39 (10.5)	**0.018**
MSCT Annulus (mm)	24.7±2.4	23.8±2.2	24.3±1.6	0.373
MSCT Annulus area (cm^2)	483.5±96.6	477.1±96.3	486.8±96.8	0.378
MSCT Annulus perimeter (mm)	77.5±7.5	77.0±7.5	77.8±7.6	0.371
MSCT Mean CFA size (mm)	7.2±0.9	6.4±0.8	7.6±1.1	**<0.001**

Data are shown as mean±standard deviation or n (%) where appropriate, SA=simplified approach, NSA=non-simplified approach, BMI=Body Mass Index, NYHA=New York Heart Association, DM=Diabetes Mellitus, HT=Hypertension, CABG=Coronary artery Bypass Grafting, MVR=Mitral Valve Replacement, COPD=Chronic Obstructive Pulmonary Disease, AF=Atrial Fibrillation, STS=Society of Thoracic Surgeons, CAD=Coronary Artery Disease, LDL=Low-density lipoprotein, HDL=High-density lipoprotein, CK-MD=Creatine kinase-MB, LVEF=Left Ventricular Ejection Fraction, AVA=Aortic Valve Area, sPAP=systolic Pulmonary Artery Pressure, MSCT=Multi-slice Computed Tomography, CFA=Common femoral artery

able Lotus valve, was placed without rapid pacing. Following valve deployment, TTE and aortography were utilized to evaluate the extent of perivalvular leak (PVL). Percutaneous closure devices (Perclose ProGlide and Prostar XL Abbott Laboratories, Abbott Park, IL) were used the majority of patients who preferred transfemoral access. Since the number of cases with surgical cutdown performed in the transfemoral access site was small, it was not excluded from the study, considering that it did not present statistical significance. Mortality, stroke, bleeding, vascular complications, device success, and adverse events were assessed due to Valve Academic Research Consortium (VARC)-2 definitions [8].

The decision to perform simplified versus non-simplified TAVI was made by experienced operators based on anatomical and clinical characteristics, including frailty, vascular access quality, comorbidities, and hemodynamic stability. While this reflects real-world practice, it may introduce selection bias due to the non-randomized nature of the study. However, to maximize standardization, institutional protocols were established for key procedural elements such as anesthesia type, use of TEE, and predilatation. These protocols were aligned with current guideline recommendations and regularly updated based on multidisciplinary team evaluations. Heart team followed a standardized workflow for pre-procedural imaging, device sizing, and post-implantation assessment, ensuring procedural consistency. Furthermore, periodic inter-site meetings were conducted to review and harmonize procedural strategies.

Statistical Analysis

All statistical analyses were applied in SPSS version 22.0. Categorical variables were reported as counts and percentages, and compared using the chi-squared test. Continuous variables were expressed as mean ± standard deviation. Depending on the distribution, comparisons between two groups were performed using either the Student's t-test or the Mann-Whitney U test. Stepwise multivariate logistic regression analysis was conducted to evaluate the association between total mortality and variables such as gender, STS score, predilatation, general anesthesia, TEE guidance, sapien 3 THV and simplified approach in the study population. The Kaplan–Meier method and the log-rank test were performed to estimate the cumulative incidences of mortality. A two-tailed P value <0.05 was considered statistically significant.

RESULTS

Of the 517 patients included in this retrospective study, TAVI was performed under SA in 144 patients and under NSA in 363 patients. There was no statistically significant difference between the groups in terms of age, gender, body mass index (BMI), functional capacity, and comorbidities except moderate to severe COPD and renal replacement therapy (44.4% vs. 41.2%; P=0.010, 4.9% vs. 1.3%; P=0.017; respectively). Therefore, no difference was found between surgical risk scores. Based on the STS score, both groups were included, intermediate-risk patients. Table 1 outlines the baseline demographics and clinical features of the study population. Baseline echocardiographic parameters were not similar between groups. The valve area was smaller (0.69±0.16 cm2 vs. 0.66±0.16 cm2; P=0.036), aortic mean gradient, and peak velocity were significantly higher in the NSA group (P<0.001). In contrast, moderate-severe mitral regurgitation was more prevelant in the SA group (17.0% vs. 10.5; P=0.018, respectively). Table 2 presents the procedural outcomes. No significant differences were observed between the groups regarding valve size, post-dilatation, or device success. Percutaneous transfemoral closure device was preferred in 99.3 % of the SA group and 97.1 % of the NSA group. Perclose ProGlide ™ Suture-Mediated Closure System was used in 80% of the SA group and 56% of the NSA group. The NSA group was treated with the most Sapien XT valve (69.4% vs. 92.8; P: <0.001). 18.8% of the valves in the SA group were S3, and 9.0% were Lotus valves.

Table 3 demonstrated the clinical outcomes of all patients. The average follow-up duration for the entire cohort was 15.7 ± 15 months. Post-procedural complications based on VARC-2 criteria - including pacemaker implantation, pericardial effusion, stroke, major vascular complications, and newly developed arrhythmias - did not differ significantly between the groups. Additionally, no significant changes were observed in laboratory parameters before and after the procedure. In-hospital mortality was 2.1% in the SA group vs. 4.3% in the NSA group (P=0.232). In the SA group,

Table 2. Procedure details, related complications and outcomes

Parameters	All (n=517)	SA group (n=144)	NSA group (n=373)	P value
Closure method, n (%)				
Prostar	177 (34.2)	20 (18.8)	157 (40.2)	**<0.001**
Proglide	328 (63.4)	116 (80.6)	212 (56.8)	
Cut-down	12 (2.3)	1 (0.7)	11 (2.9)	
Valve size (mm), n (%)				
20	1 (0.2)	-	1 (0.3)	0.067
23	215 (41.7)	61 (42.4)	154 (41.4)	
25	14 (2.7)	6 (4.2)	8 (2.2)	
26	213 (41.3)	52 (36.1)	161 (43.3)	
27	5 (1.0)	4 (2.8)	1 (0.3)	
29	68 (13.2)	21 (14.6)	47 (12.6)	
Edwards SAPIEN XT, n (%)	446 (86.3)	100 (69.4)	346 (92.8)	**<0.001**
Edwards SAPIEN 3, n (%)	41 (7.9)	27 (18.8)	14 (3.8)	**<0.001**
LOTUS, n (%)	22 (4.3)	13 (9.0)	9 (2.4)	**0.002**
Predilatation, n (%)	373 (72.1)	0.0	373 (100.0)	**<0.001**
Postdilatation, (%)	17 (3.3)	5 (3.5)	12 (3.2)	0.874
PostTAVI creatinine (mg/dL)	0.97±0.40	1.01±0.39	0.95±0.40	0.145
PostTAVI hemoglobin (mg/dL)	10.6±1.7	10.6±1.7	10.4±2.1	0.308
PostTAVI troponin (pg/mL)	309.1±812.1	309.1±812.1	212.8±431.0	0.122
PostTAVI CK-MB (ng/mL)	7.5±5.9	7.5±5.9	14.3±98.8	0.591
Device success, n (%)	501 (96.9)	141 (97.9)	360 (96.5)	0.409

Data are shown as mean±standard deviation or n (%) where appropriate, SA=simplified approach, NSA=non-simplified approach, TAVI=Transcatheter Aortic Valve Implantation, CK-MD=Creatine kinase-MB,

there was no 30-day mortality. Although there was a trend toward lower mortality at 30-day, it was not statistically significant (0% vs. 2.9%, respectively, P=0.058), while the rate of in-hospital, six months, and 1-year mortality was similar. However, total cumulative mortality at the end of the follow-up period was statistically significantly lower in the SA group (7.6% vs. 35.7; P<0.001) (Fig. 1).

Both groups demonstrated favorable bioprosthetic valve performance, with comparable final mean gradients. Follow-up echocardiographic parameters are presented in Table 4, with no significant differences observed between the groups when PVL was compared on the post-TAVI, 30-day, and 1st year. The increase in the AVA and the decrease in mean aortic gradient continued steadily for 30-day and 1st year.

Multivariate logistic regression analyses were performed to identify independent predictors of total mortality following the TAVI procedure. The analysis revealed that gender, STS score, predilatation, use of general anesthesia, TEE guidance, and the simplified approach independently predicted overall mortality (Table 5). The use of Sapien 3 THV, which differed between the two groups, was not one of the independent predictors of total mortality. Fig. 2 illustrates the Kaplan-Meier survival curves for patients in the SA and NSA groups. A statistically significant difference in overall survival probability was observed between the two groups. (Overall: 41.1±1.8 month, 95%CI: 37.6-44.7; P:0.019; SA Group: 43.7±1.3 month, 95% CI: 41.1-46.4 month; NSA Group: 40.1±1.8 month; 95% CI: 36.4-43.7).

Table 3. Follow up outcomes

Parameters	All (n=517)	SA group (n=144)	NSA group (n=373)	P value
30-day NYHA, n (%)				
1	131 (41.8)	95 (46.1)	36 (40.2)	0.393
2	162 (50.9)	44 (49.4)	118 (51.5)	
3	23 (7.2)	4 (4.5)	19 (8.3)	
Pacemaker, n (%)	39 (7.5)	11 (7.6)	28 (7.5)	0.959
Stroke, n (%)	2 (0.4)	0.0	2 (0.5)	0.379
Pericardial effusion, n (%)	9 (1.8)	1 (0.7)	8 (2.1)	0.488
Emerging arrhythmia n (%)				
AF	19 (3.7)	5 (3.5)	14 (3.8)	0.499
VT	3 (0.6)	0 (0.0)	3 (0.8)	
LBBB	13 (2.5)	6 (4.2)	7 (1.9)	
Major vascular complication, n (%)	32 (6.1)	12 (8.4)	20 (5.4)	0.180
Closure device failure, n (%)	11.0 (2.1)	3 (2.1)	8 (2.1)	0.987
Acute renal failure, n (%)	4 (0.8)	2 (1.4)	2 (0.5)	0.587
Discharge time (day)	4.5±2.4	4.6±2.4	4.5±2.2	0.740
In-hospital mortality, n (%)	19 (3.7)	3 (2.1)	16 (4.3)	0.232
30-day mortality, n (%)	10 (2.1)	0.0	10 (2.9)	0.058
6th month mortality, n (%)	7 (1.7)	2 (2.4)	5 (1.5)	0.590
1st year mortality, n (%)	48 (12.1)	5 (6.6)	43 (13.4)	0.100
Total mortality, n (%)	144 (27.9)	11 (7.6)	133 (35.7)	**<0.001**

Data are shown as n (%). SA=simplified approach, NSA=non-simplified approach, AF=Atrial Fibrillation, LBBB=Left Bundle Branch Block, VT=Ventricular tachycardia, NYHA=New York Heart Association

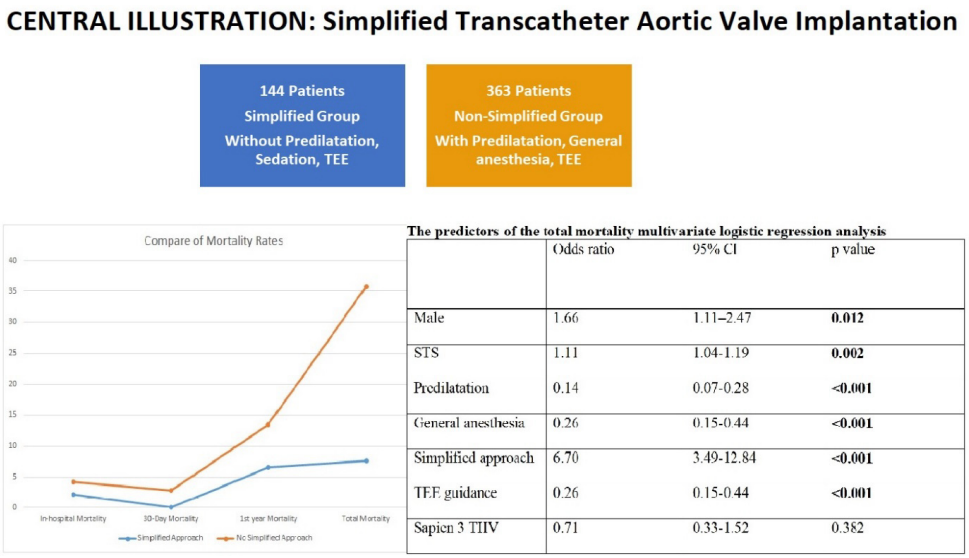

Fig. 1. Although there was no significant difference in mortality in post-procedural, first-month, and first-year follow-up, a significant difference was found in favor of the Simplified group in total mortality.

Table 4. Follow up echocardiographic parameters

Parameters	All (n=517)	SA group (n=144)	NSA group (n=373)	P value
PostTAVI LVEF (%)	54.2±12.8	54.5±12.8	54.1±12.8	0.770
PostTAVI aortic mean gradient (mm Hg)	10.3±3.6	10.5±3.8	10.3±3.6	0.593
PostTAVI sPAP (mmHg)	36.9±13.0	37.5±14.7	36.7±12.3	0.536
PostTAVI PVL, n (%)				
Mild	92 (18.5)	17 (12.6)	75 (20.8)	0.238
Moderate	5 (1.0)	2 (1.4)	3 (0.8)	
30-day LVEF (%)	55.4±11.4	55.2±11.2	55.4±11.5	0.862
30-day Aortic mean gradient (mm Hg)	10.9±4.3	10.9±3.8	10.8±4.5	0.908
30-day sPAP (mmHg)	36.7±13.8	35.4±15.1	37.2±13.3	0.329
30-day PVL, n (%)				
Mild	51 (17.5)	12 (14.5)	39 (18.7)	0.735
Moderate	6 (2.1)	2 (2.4)	4 (1.9)	
1st year LVEF (%)	58.5±8.7	58.6±8.6	58.5±8.8	0.978
1st year aortic mean gradient (mm Hg)	12.2±4.4	12.7±5.1	12.1±4.4	0.688
1st year sPAP (mmHg)	35.9±14.0	38.0±16.0	35.6±13.8	0.596

Data are shown as mean±standard deviation or n (%) where appropriate. TAVI=Transcatheter Aortic Valve Implantation, SA=simplified approach, NSA=non-simplified approach, LVEF=Left Ventricular Ejection Fraction, sPAP=systolic Pulmonary Artery Pressure, PVL=Paravalvular Leakage

DISCUSSION

In this retrospective study with the largest simplified approach group, we aimed to demonstrate the safety and feasibility of TAVI without predilatation, general anesthesia, and TEE guidance. Our research found significant differences between SA and no NSA groups.

The major findings are: (1) there were no differences in baseline demographic characteristics and comorbidities between the two groups, but aortic valve stenosis was more severe in the NSA group. (2) The device success rates were high in both groups, and similar peri-procedural and mid-term major complication rates were found. (3) There was no difference between the short and mid-term performance of the transcatheter valve between the two groups. (4) Al-

Table 5. The predictors of the total mortality multivariate logistic regression analysis

	Odds ratio	95% CI	P value
Male	1.66	1.11–2.47	**0.012**
STS	1.11	1.04-1.19	**0.002**
Predilatation	0.14	0.07-0.28	**<0.001**
General anesthesia	0.26	0.15-0.44	**<0.001**
Simplified approach	6.70	3.49-12.84	**<0.001**
TEE guidance	0.26	0.15-0.44	**<0.001**
Sapien 3 THV	0.71	0.33-1.52	0.382

STS=Society of Thoracic Surgeons, TEE=Transesophageal Echocardiography, THV=Transcatheter Heart Valve, CI=Confidence Interval

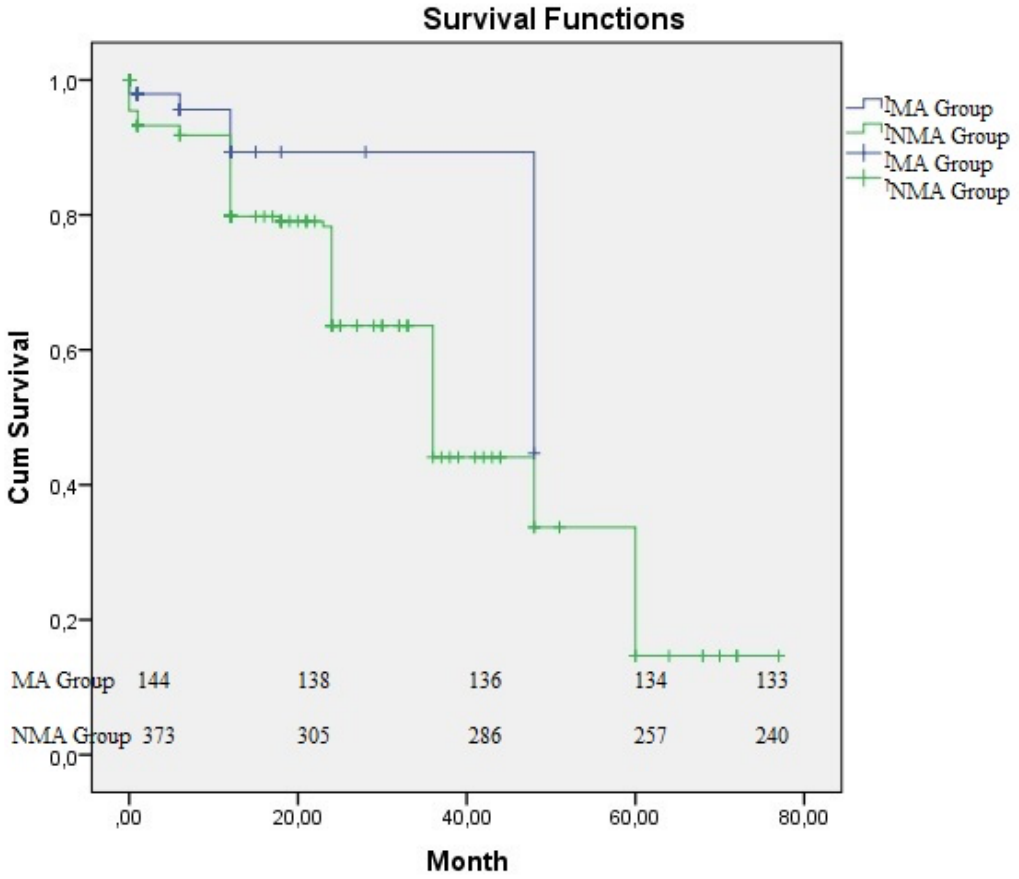

Fig. 2. Kaplan-Meier analysis of survival curves in patients with all Simplified and No Simplified groups. Overall survival probability was significantly different in those patients (Overall: 41.1±1.8 month, 95%CI: 37.6-44.7; P=0.019; SA Group: 43.7±1.3 month, 95% CI: 41.1-46.4 month,; NSA Group: 40.1±1.8 month; 95% CI: 36.4-43.7).

though there was no difference at mortality rates post TAVI, 30-day, 1st-year mortality between the two groups, total cumulative mortality at the end of the follow-up period was lower in the SA group, and the simplified approach was found to be an independent risk factor for total mortality.

With the rapid increase in the TAVI procedure, the nature of the procedure has transformed, becoming less surgical and more comparable to PCI. There is a trend around the world towards the minimalist approach in the pre-procedural, procedural, and post-procedural steps [9]. To minimize the procedure during the TAVI, we switch from general anesthesia to conscious anesthesia, from surgical cut-down to percutaneous access, from contralateral femoral access to radial access, from predilatation to no predilatation.

According to the results of the UK TAVI registry [10], between 2007 and 2014, especially at experienced centers, predilatation decreased by up to 50%. The no predilatation TAVI procedure was first evaluated by Grübe *et al*. [11] in 2011, and its safety and feasibility were assessed on non-randomized retrospective studies and meta-analyses. One of the most critical advantages of predilatation is that it reduces the device malposition in patients with inadequate anatomical features and thus eliminates the need for a second device. Kim *et al*. [12] reported device malposition in two patients with severe calcific valve, which showed us the importance of patient selection before deciding no-predilatation. Similarly, Chan *et al*. [13] reported two cases where predilatation was not initially planned but was ultimately required to facilitate device crossing and implantation. In contrast to other studies [12, 14], similar to Spaziano *et al*. [15] previously reported, device success rates were similar and high between the two groups in our study. These find-

ings may be due to the reduced crossing profile of S3-THV and the predilatation of patients with very complex anatomies or highly calcified valves. In the case of under expansion, post dilatation can be performed readily. In our study, there was no difference in the number of post dilatation between the two groups, and no patient had severe PVL. In our clinical practice, the decision of predilatation was made by an experienced operator based on the bicuspid aortic valve, horizontal aorta, asymmetric valve configuration, disproportionate excess calcification in native non-coronary cuspis. Another situation where predilation may be advantageous is when the exact valve size cannot be determined or in the case of low coronary origins. But since the pre-procedural multimodality approach has become widespread, this benefit has now become historical.

Predilatation has many damaging severe effects; fracture and embolization of calcific nodules by balloon valvuloplasty, separating fused commissures, and distortion of the native aortic valve ring. Thus, stroke and annular rupture are one of the feared and catastrophic consequences of predilatation. Also, overstretching of the annulus with balloon dilatation may disrupt valve sizing and cause post-procedural PVL [16]. The second most crucial disadvantage is that hemodynamic deterioration with rapid pacing, especially in patients with reduced systolic function, leads to worse outcomes in this fragile patient group. Secondary to rapid pacing, bradycardia, AV block, and severe aortic insufficiency may occur and may require positive inotropic support. However, there are conflicting results in studies related to rapid pacing. In the EASE-IT transfemoral multicenter registry [17], this rate was very low (1.0% overall), while it was found to be quite high in the EASE-IT transapical recording study (22.2%) [18]. Third, predilatation to the severe bulky calcific native aortic valve may increase the risk of coronary embolism, pre-procedural myocardial infarction, and stroke. However, Bijuklic *et al*. [19] suggested that the incidence and number of cerebral lesions were numerically higher, though not statistically significant, in patients undergoing TAVI without predilatation. Similarly, in another study with self-expandable valve implantation, 30-day stroke rates were higher in the predilatation group [20]. But this finding is thought to be irrational. The homogeneous disintegration of calcific debris can only explain this situation during predilatation. But similar to the results of many studies, stroke rates were non-significantly different between the two groups in our study. Finally, predilatation, especially with large-sized balloons, can lead to conduction disturbances and the need for permanent pacemakers. Lange *et al*. showed that predilatation with small valvuloplasty balloon in CoreValve implanted patients reduced permanent pacemaker rates [21]. In contrast, pacemaker rates were similar between the two groups in our study. In the abovementioned UK TAVI registry, which has the most significant number of patients (n=5,887), patients with and without predilatation were evaluated. Consistent with our findings, the skipping predilatation step did not correlate with an increased incidence of short-term adverse outcomes, particularly when the balloon-expandable SAPIEN (Edwards Lifesciences Inc., Irvine, CA) valve [10]. At most recently published prospective EASE-IT multicentre registry17 transfemoral balloon-expandable THV was implanted in 196 patients, 56 of which underwent predilatation, and 140 without predilatation. The EASE-IT TF registry demonstrated that omitting predilatation in transfemoral TAVI using balloon-expandable valves did not adversely affect short-term clinical outcomes, aligning with our observation that predilatation was not an independent predictor of procedural success or mortality. However, omitting predilatation is linked to shorter procedural times, which may be advantageous in selected patient populations.

Although there is a tendency to perform TAVI under general anesthesia and with TEE, until the learning curve is complete, the first step of the simplified approach is to perform TAVI under conscious sedation and without TEE guidance. Albeit there is no difference in the duration of hospital stay between the two groups in our study, there are disadvantages such as prolonging the procedure time, cost, and length of hospital stay of general anesthesia. The fact that general anesthesia is an independent predictor of total mortality may be related to anesthesia-related pulmonary infections, extubation difficulty, respiratory distress, catecholamine requirement, and transfusion need in this fragile patient group [22].

Fava *et al*. [23] evaluated the impact of the minimalist approach (MA) on clinical outcomes of the TAVI procedure. To investigate this, the authors analyzed 229 consecutive patients who underwent mini-

malist approach TAVI (MA-TAVI), compared to 74 patients treated with the standard approach. Their findings confirmed that MA-TAVI is both feasible and safe, resulting in shorter procedure times and reduced hospital stays, while maintaining comparable 30-day outcomes and offering improved patient comfort [23]. Babaliaros *et al*. [24] reported the first analysis of mid-term mortality outcomes and associated costs of a minimalist approach to TAVI in the United States. In the study in which 142 patients were evaluated (MA-TF TAVI, n=70 and standart approach-TF TAVI, n=72), it was shown that MA-TF resulted in lower costs due to a shorter hospital stay and limited resource use. Moreover, 30-day mortality did not differ significantly between the MA-TF TAVR group and the standard approach group (0% vs. 6%, P=0.12), and the incidence of stroke or transient ischemic attack at 30 days was also comparable (4.3% vs. 1.4%, P=0.35). Rates of moderate or severe paravalvular leak and device success at 30 days were similar between the MA-TF TAVR and standard TF-TAVR groups (3% vs. 5.8%, P=0.4 and 90% vs. 88%, P=0.79, respectively).In the FAST trial published by Lefèvre *et al*. [25] reviewed all TF-TAVR retrospectively. The FAST strategy compromise of local anesthesia with conscious sedation, ultrasonographic guided access site puncture, radial approach for secondary arterial access, and left ventricular guidewire rapid pacing. In a study comparing 76 patients treated with the FAST protocol to 209 patients managed with the standard approach, the primary endpoint - an early safety composite including all-cause mortality, stroke, life-threatening bleeding, acute kidney injury, coronary obstruction, major vascular complications, and valve-related dysfunction - was significantly lower in the FAST group (1.3% vs. 14.3%; P<0.001). The FAST protocol was also associated with reductions in major bleeding (1.3% vs. 10.1%; P=0.01), blood transfusion requirements (2.6% vs. 14.3%; P<0.01), and vascular complications related to secondary access (0.0% vs. 5.3%; P=0.04). The primary aim of this study was to assess the feasibility and safety of the minimally invasive FAST protocol compared to the standard approach. Building upon these trials, our study contributes new insights by including a larger and more diverse patient population, with both simplified and non-simplified approaches evaluated in parallel. This allowed for a robust comparison of procedural elements and long-term mortality, reinforcing the clinical value of simplification strategies in real-world practice.

Interestingly, in the multivariate logistic regression analysis, variables such as TEE guidance, general anesthesia, and predilatation were associated with lower odds ratios (OR<1) for total mortality, suggesting a potential protective effect. However, these practices were not utilized in the simplified approach group, which paradoxically demonstrated significantly lower overall mortality. This apparent discrepancy may be explained by the collinearity between these procedural variables and the broader clinical context in which they were applied. Specifically, general anesthesia, TEE guidance, and predilatation were often reserved for more complex or high-risk anatomies, potentially introducing confounding by indication. Furthermore, while these individual elements may appear protective in isolation, their combined omission within a structured and streamlined simplified protocol - alongside careful patient selection and standardized peri-procedural management - may offset the need for such interventions and ultimately contribute to improved outcomes.

Limitations

In our study compared to the NSA group, SA-TAVI had statistically significant lower total mortality rates. Although there is no statistical difference in post-TAVI pacemaker need, stroke and PVL rates between the two groups, the positive synergistic effect of avoiding general anesthesia, TEE guidance, and predilatation may have caused a reduction in total mortality. Because of the retrospective nature of our study, patients could not be randomized. This could be considered a limitation of this research. In our center, the decision of predilatation and balloon size was left to the conclusion of an experienced operator.

CONCLUSION

In conclusion, given the results of our study, we showed the feasibility and safety of SA-TAVI independent of anatomic selection criteria on an all-comers. It was safe to skip the predilatation step, transition to conscious sedation and not using routine TEE guidance during the TAVI procedure, and wasn't related to adverse events. Besides this, the implantation of the

transcatheter aortic valve with a simplified approach was associated with lower total mortality.

Ethics Approval and Consent to Participate

This study was approved by the Gazi University Rectorate's Office Ethics Commission (Decision No: 2025-1207/11; date: 08.07.2025). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Written informed consent was obtained from all individual participants included in the study.

Data Availability

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

Authors' Contribution

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

Conflict of Interest

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

Financing

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

Acknowledgments

The author extends profound gratitude to Doç.Dr.Burak Sezenöz for his esteemed guidance and contributions to this study

Generative Artificial Intelligence Statement

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

Editor's Note

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

REFERENCES

1. Cribier A, Eltchaninoff H, Bash A, et al. Percutaneous transcatheter implantation of an aortic valve prosthesis for calcific aortic stenosis: first human case description. Circulation. 2002;106(24):3006-3008. doi: 10.1161/01.cir.0000047200.36165.b8.
2. Kodali SK, Williams MR, Smith CR, kar RR, et al; PARTNER Trial Investigators. Two-year outcomes after transcatheter or surgical aortic-valve replacement. N Engl J Med. 2012;366(18):1686-1695. doi: 10.1056/NEJMoa1200384.
3. Makkar RR, Fontana GP, Jilaihawi H, et al; PARTNER Trial Investigators. Transcatheter aortic-valve replacement for inoperable severe aortic stenosis. N Engl J Med. 2012;366(18):1696-1704. doi: 10.1056/NEJMoa1202277.
4. Popma JJ, Deeb GM, Yakubov SJ, et al; Evolut Low Risk Trial Investigators. Transcatheter Aortic-Valve Replacement with a Self-Expanding Valve in Low-Risk Patients. N Engl J Med. 2019;380(18):1706-1715. doi: 10.1056/NEJMoa1816885.
5. Akodad M, Lefèvre T. TAVI: Simplification Is the Ultimate Sophistication. Front Cardiovasc Med. 2018;5:96. doi: 10.3389/fcvm.2018.00096.
6. Pagnesi M, Baldetti L, Del Sole P, et al. Predilatation Prior to Transcatheter Aortic Valve Implantation: Is it Still a Prerequisite? Interv Cardiol. 2017;12(2):116-125. doi: 10.15420/icr.2017:17:2.
7. Selle A, Figulla HR, Ferrari M, et al. Impact of rapid ventricular pacing during TAVI on microvascular tissue perfusion. Clin Res Cardiol. 2014;103(11):902-911. doi: 10.1007/s00392-014-0728-9.
8. Kappetein AP, Head SJ, Généreux P, et al. Updated standardized endpoint definitions for transcatheter aortic valve implantation: the Valve Academic Research Consortium-2 consensus document. Eur Heart J. 2012;33(19):2403-2418. doi: 10.1093/eurheartj/ehs255.
9. Wood DA, Lauck SB, Cairns JA, et al. The Vancouver 3M (Multidisciplinary, Multimodality, But Minimalist) Clinical Pathway Facilitates Safe Next-Day Discharge Home at Low-, Medium-, and High-Volume Transfemoral Transcatheter Aortic Valve Replacement Centers: The 3M TAVR Study. JACC Cardiovasc Interv. 2019;12(5):459-469. doi: 10.1016/j.jcin.2018.12.020.
10. Martin GP, Sperrin M, Bagur R, et al. Pre-Implantation Balloon Aortic Valvuloplasty and Clinical Outcomes Following Transcatheter Aortic Valve Implantation: A Propensity Score Analysis of the UK Registry. J Am Heart Assoc. 2017;6(2):e004695. doi: 10.1161/JAHA.116.004695.
11. Grube E, Naber C, Abizaid A, et al. Feasibility of transcatheter aortic valve implantation without balloon pre-dilation:

a pilot study. JACC Cardiovasc Interv. 2011;4(7):751-757. doi: 10.1016/j.jcin.2011.03.015.
12. Kim WK, Praz F, Blumenstein J, et al. Transfemoral aortic valve implantation of Edwards SAPIEN 3 without predilatation. Catheter Cardiovasc Interv. 2017;89(1):E38-E43. doi: 10.1002/ccd.26464.
13. Chan PH, Mario CD, Moat N. Transcatheter aortic valve implantation without balloon predilatation: not always feasible. Catheter Cardiovasc Interv. 2013;82(2):328-332. doi: 10.1002/ccd.24381.
14. Möllmann H, Kim WK, Kempfert J, et al. Transfemoral aortic valve implantation of Edwards SAPIEN XT without predilatation is feasible. Clin Cardiol. 2014;37(11):667-671. doi: 10.1002/clc.22318.
15. Spaziano M, Sawaya F, Chevalier B, et al. Comparison of Systematic Predilation, Selective Predilation, and Direct Transcatheter Aortic Valve Implantation With the SAPIEN S3 Valve. Can J Cardiol. 2017;33(2):260-268. doi: 10.1016/j.cjca.2016.09.007.
16. Kempfert J, Meyer A, Kim WK, et al. First experience without pre-ballooning in transapical aortic valve implantation: a propensity score-matched analysis. Eur J Cardiothorac Surg. 2015;47(1):31-38; discussion 38. doi: 10.1093/ejcts/ezu097.
17. Schymik G, Rudolph T, Jacobshagen C, et al. Balloon-expandable transfemoral transcatheter aortic valve implantation with or without predilation: findings from the prospective EASE-IT TF multicentre registry. Open Heart. 2019;6(2):e001082. doi: 10.1136/openhrt-2019-001082.
18. Strauch J, Wendt D, Diegeler A, et al. Balloon-expandable transapical transcatheter aortic valve implantation with or without predilation of the aortic valve: results of a multicentre registry. Eur J Cardiothorac Surg. 2018;53(4):771-777. doi: 10.1093/ejcts/ezx397.
19. Bijuklic K, Haselbach T, Witt J, et al. Increased Risk of Cerebral Embolization After Implantation of a Balloon-Expandable Aortic Valve Without Prior Balloon Valvuloplasty. JACC Cardiovasc Interv. 2015;8(12):1608-1613. doi: 10.1016/j.jcin.2015.07.013.
20. Pagnesi M, Jabbour RJ, Latib A, et al. Usefulness of Predilation Before Transcatheter Aortic Valve Implantation. Am J Cardiol. 2016;118(1):107-112. doi: 10.1016/j.amjcard.2016.04.018.
21. Lange P, Greif M, Vogel A, et al. Reduction of pacemaker implantation rates after CoreValve® implantation by moderate predilatation. EuroIntervention. 2014;9(10):1151-1157. doi: 10.4244/EIJV9I10A195.
22. Barbanti M, Gulino S, Costa G, Tamburino C. Optimization and simplification of transcatheter aortic valve implantation therapy. Expert Rev Cardiovasc Ther. 2018;16(4):287-296. doi: 10.1080/14779072.2018.1449644.
23. Fava C, Gamboa P, Caponi G, G.mez C, Salmo F, Guevara E, et al. Minimalist Approach for Percutaneous Aortic Valve Implantation. Rev Argent Cardiol 2019;87(4):265-270. doi: 10.7775/rac.v87.i4.12015.
24. Babaliaros V, Devireddy C, Lerakis S, et al. Comparison of transfemoral transcatheter aortic valve replacement performed in the catheterization laboratory (minimalist approach) versus hybrid operating room (standard approach): outcomes and cost analysis. JACC Cardiovasc Interv. 2014;7(8):898-904. doi: 10.1016/j.jcin.2014.04.005.
25. Lefèvre G, Jégou A, Dambrin G, et al. Comparison of Trans-Femoral Transcatheter Aortic Valve Replacement Performed With a Minimally Invasive Simplified Technique: "FAST" Versus a Standard Approach. J Invasive Cardiol. 2019;31(10):300-306.
26. Fiorina C, Maffeo D, Curello S, et al. Direct transcatheter aortic valve implantation with self-expandable bioprosthesis: feasibility and safety. Cardiovasc Revasc Med. 2014;15(4):200-203. doi: 10.1016/j.carrev.2014.03.005.

The European Research Journal 2025;11(5):989-1000 Original Article

DOI: https://doi.org/10.18621/eurj.1742863 *Psychology*

Attachment representations in individuals diagnosed with schizophrenia: A projective study through the bird's nest drawing

Volkan Demir

Department of Psychology, İstanbul Gelişim University, Faculty of Economics, Administrative and Social Sciences, İstanbul, Türkiye

ABSTRACT

Objectives: This study aims to explore the attachment representations of individuals diagnosed with schizophrenia through the Bird's Nest Drawing, a projective technique. It also seeks to compare the findings with those of individuals without a schizophrenia diagnosis to identify potential differences in attachment-related imagery and narrative themes.

Methods: A total of 100 participants took part in the study: 50 individuals diagnosed with schizophrenia and 50 without any psychiatric diagnosis. Each participant was asked to draw a bird's nest and write a brief narrative about their drawing. Quantitative data were analyzed using chi-square tests. Qualitative analysis was conducted through thematic evaluation of narratives from 15 randomly selected participants from each group.

Results: Quantitative analysis showed statistically significant group differences in several aspects of the drawings, including the presence of parent bird and chick/egg, use of appropriately colored figures, frequent use of green, and whether the nest touched a surface. Qualitative analysis revealed that narratives of individuals with schizophrenia included themes of loneliness, mistrust, unmet basic needs, and disconnection from the nest. Conversely, narratives of individuals without schizophrenia reflected familial closeness, caregiving, and emotional security.

Conclusions: The results suggest that individuals with schizophrenia project attachment-related difficulties - such as distrust, disconnection, and impaired bonding - more prominently in both their drawings and written narratives. The BND test appears to be a meaningful tool for exploring attachment dynamics at the projective level in clinical populations.

Keywords: Schizophrenia, attachment, projective drawing, bird's nest drawing

Attachment theory posits that the emotional bonds formed between an individual and their primary caregiver during early childhood leave lasting impacts on interpersonal relationships, emotion regulation capacity, and self-perception throughout life [1, 2]. According to this perspective, secure attachment developed in early childhood contributes to healthier social relationships, more effective coping with stress, and greater psychological resilience [3]. On the other hand, insecure or avoidant attachment

Received: July 15, 2025 **Accepted:** August 24, 2025 **Available Online:** August 26, 2025 **Published:** September 4, 2025

How to cite this article: Demir V.Attachment representations in individuals diagnosed with schizophrenia: A projective study through the bird's nest drawing. Eur Res J. 2025;11(5):989-1000. doi: 10.18621/eurj.1742863

Corresponding author: Volkan Demir, MD., Assist. Prof., Phone: +90 212 422 70 00, E-mail: vdemir@gelisim.edu.tr

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

patterns can lead to a decrease in the individual's basic sense of trust, difficulties in social interactions, and impaired psychological functioning [4]. Recent studies have shown that attachment styles are an important factor that affects not only social adjustment but also the development of psychopathological processes [5, 6].

Although Bowlby [2] originally developed attachment theory to explain childhood development, it has since evolved into a valuable framework for understanding adult psychopathology as well [7, 8]. Attachment representations are known to influence various psychological functions, particularly emotional regulation, self-image, and interpersonal relationships. Disruptions in these areas may contribute to the onset of psychiatric disorders. Among these, schizophrenia is considered a complex and heterogeneous mental condition, believed to result from abnormalities in brain development shaped by the interplay of genetic predispositions and psychosocial influences [9]. The disorder is commonly characterized by impaired social functioning; individuals may experience difficulty in engaging with others or maintaining social support [10]. Interestingly, researchers emphasize that attachment styles not only influence how psychological disorders begin, but also how they progress over time [11]. It is believed that these attachment dynamics affect both the severity of symptoms and the individual's level of daily functioning [8]. In this context, the possible links between attachment process disruptions and psychotic disorders such as schizophrenia are increasingly being investigated. Indeed, numerous studies have shown that individuals diagnosed with schizophrenia often exhibit insecure attachment styles [7, 12].

Evaluating attachment styles in individuals diagnosed with schizophrenia provides a more comprehensive understanding of psychopathological symptoms and contributes to the development of tailored intervention plans [7]. Assessment tools used in this context range from self-report measures that evaluate insight into past attachment experiences and relational patterns to semi-structured clinical interviews that address deeper psychodynamic structures [5]. However, in cases such as schizophrenia, where disturbances in thought processes and reality testing are observed, individuals may have limitations in verbal expression. Thus, projective assessment methods gain importance as complementary tools for understanding attachment representations [13].

One such projective assessment tool is the Bird's Nest Drawing (BND) test, which enables individuals to symbolically and creatively express emotional traces related to early caregiver-child relationships [14]. In this task, individuals are asked to first draw a bird's nest and then write a brief story about their drawing. The number of birds, contents of the nest, its location, environmental features, and overall organization offer projective cues regarding the individual's sense of emotional security and their relational representations of caregiving figures [14, 15]. Because the test includes both visual and narrative elements, it allows for the simultaneous assessment of verbal and visual expression while providing rich material for psychodynamic interpretations based on attachment theory. The nest - symbolizing safety, shelter, protection, and relational bonding—facilitates the symbolic reflection of unconscious dynamics related to the individual's attachment system [16, 17].

A review of the existing literature reveals that most studies examining attachment representations in individuals with schizophrenia rely primarily on self-report measures. This has led to an overly surface-level and verbal exploration of attachment processes. However, the cognitive impairments, attention difficulties, and limited insight commonly observed in schizophrenia may pose challenges and limitations in administering such instruments. In an effort to overcome these limitations and explore attachment representations in a more in-depth manner, the present study employed a projective technique - the BND test.

In this study, both quantitative and qualitative data derived from the BND test were analyzed, and the following hypotheses were tested:

H1: The frequency of visual elements in the BND test that indicate attachment representations (e.g., presence of chick/parent figures, contact with a surface, appropriate use of color) will differ significantly between individuals diagnosed with schizophrenia and those without a diagnosis.

H2: The thematic content emerging in participants' narratives about their BNDs will reflect their attachment representations, and there will be significant differences in these themes between individuals with and without a diagnosis of schizophrenia.

METHODS

Research Design

This study was built on a qualitative framework, embracing a multi-method approach that brought together different ways of gathering insight. In this process, both the visual expressions found in participants' drawings and the words they used to describe them were carefully examined. Drawing on diverse data collection techniques to explore a shared theoretical focus was seen as a meaningful way to strengthen the credibility and depth of the findings [18]. Moreover, the analysis did not rely on a single perspective; instead, it was enriched through the inclusion of multiple data sources, allowing the research to rest on a more grounded and trustworthy foundation [19]. Although the study incorporates both qualitative and quantitative analyses, the quantitative data were derived entirely from the categorical coding of qualitative material obtained through the BND test and accompanying narratives. Therefore, the methodological orientation of the research remains within a qualitative framework, with quantitative analyses serving to complement and further substantiate the qualitative findings.

Participants

The study sample consisted of a total of 100 participants, including 50 individuals diagnosed with schizophrenia and 50 individuals with no psychiatric diagnosis. The ages of the schizophrenia group ranged from 23 to 49 years, with 31 men and 19 women. The comparison group, composed of individuals without a diagnosis of schizophrenia, ranged in age from 21 to 46 years, with 29 women and 21 men. Participants in the schizophrenia group were recruited from volunteers receiving inpatient treatment at a private care center in Istanbul. In terms of clinical characteristics, all individuals in the schizophrenia group had a chronic diagnosis of schizophrenia and were undergoing antipsychotic treatment. Most were receiving atypical antipsychotics, while some were prescribed typical agents or combination therapy. The duration of illness varied: 12 participants had been diagnosed for 5 years or less, 15 for 6-10 years, 13 for 11-15 years, and 10 for more than 15 years. Regarding education, 18 participants had completed primary school, 20 had completed middle school, 9 had finished high school, and 3 were university graduates. In terms of marital status, 40 were single, 8 were married, and 2 were divorced. Individuals in the non-schizophrenia group were selected from adults with no history of psychiatric illness. To determine whether the sample size was statistically adequate, a power analysis was conducted using the G*Power 3.1 program. In social sciences, a statistical power of 80% ($1-\beta = 0.80$) is generally considered sufficient. Based on this criterion, with a medium effect size (Cohen's $w = 0.30$) and a significance level of $\alpha = 0.05$, the analysis indicated that a minimum of 87 participants would be required. Thus, the inclusion of 100 participants (50 with schizophrenia and 50 without psychiatric diagnoses) ensured sufficient statistical power for the study.

Data Collection Procedure

Before beginning the study, the director of the Private Karanfil Care Center in Istanbul - where the research would take place - was consulted, and written permission was kindly granted to carry out the study at the institution. Following this approval, all required documents were prepared and submitted to the Ethics Committee of Istanbul Kent University. Ethical clearance was formally granted with the decision numbered 2025/06, dated 01.07.2025. Participants who voluntarily agreed to be part of the study were individually interviewed. Each was informed in detail about the study's purpose, the procedures involved, and how data would be collected. Their verbal and written consent was respectfully obtained before proceeding. During the session, every participant was offered an A4-sized sheet of paper and a set of colored pencils. They were gently encouraged with the following instruction: "I'd like you to draw a bird's nest. Please feel free to use any colors you wish." After they completed their drawings, they were asked softly, "Is your drawing finished? Would you like to add anything else?" Once the drawing was finalized, participants were then invited to write a short story describing the nest they had drawn. All drawings were later reviewed by the lead researcher and two clinical psychologists. The researcher holds a doctorate in psychology and has conducted multiple studies in the areas of art therapy and projective drawing assessments. The other two professionals are trained clinical psychologists with established experience in art-based therapeutic evaluations.

Data Collection Tools

In order to explore attachment representations through projective means, two main tools were used in this study: a Demographic Information Form and the BND test.

Demographic Information Form

It was designed by the researcher to collect data on the basic sociodemographic characteristics of the participants. The form includes questions about the participants' identifying information, such as gender and age.

Bird's Nest Drawing (BND) Test

The BNDTest, first introduced by Kaiser in 1996 and later adapted for Turkish populations by Demirbağ (2016), is a projective technique aimed at exploring individuals' implicit attachment patterns through visual symbolism. In its administration, participants receive a blank A4 sheet and colored pencils, and are simply invited to "draw a bird's nest," with no additional guidance. This open-ended approach allows for the free emergence of symbolic themes related to attachment experiences.

For consistent evaluation, the Turkish adaptation applies the BND Assessment Scale, which consists of 12 dichotomous items scored as 1 ("present") or 0 ("absent"). These items capture both the compositional and symbolic qualities of the drawing, such as:

1. Presence of a parent bird in or near the nest
2. Presence of a baby bird or eggs in or near the nest
3. Use of green as the dominant color
4. Utilization of more than 20% of the total page area
5. Inclusion of a sun symbol
6. Inclusion of a cloud symbol
7. Presence of "M"-shaped birds in the sky
8. Presence of additional decorative or contextual elements (e.g., flowers, pets)
9. Use of realistic colors for depicted objects
10. Nest illustrated as trapped inside the tree trunk (reverse scored)
11. Nest placed on a visible supporting surface (e.g., branch, leaf, ground)
12. Presence of both a parent bird and a baby bird or egg.

Higher scores on the scale are interpreted as reflecting more secure attachment representations, whereas the reverse-scored item (nest trapped in the trunk) is associated with symbolic restriction or confinement. The rubric was developed through comparative analysis of drawings from individuals with differing attachment styles. Reliability analyses indicated inter-rater agreement ranging from .66 to 1.00 for individual items, with overall internal consistency coefficients of .70 (KR-20) and .73 (split-half). Criterion validity was supported by a statistically significant, moderate positive correlation with Kern's Secure Attachment Scale ($r = .34$, $P<0.01$).

Statistical Analysis

In the qualitative phase of the study, participants' BNDs and the narratives they produced to accompany them were examined through a thematic analysis process. The analysis began with familiarization with the data, during which the drawings and narratives were reviewed multiple times to capture initial impressions and recurring elements. Coding was then carried out, distinguishing between two main categories: predetermined themes derived from existing literature (e.g., security, nurture, isolation) and emergent themes that arose inductively from the participants' own expressions. The thematic framework was refined through an iterative process of reviewing and grouping related codes, followed by naming and defining the themes. The drawings and narratives were independently evaluated by the lead researcher and two domain experts, after which evaluations were compared to identify consensus and discrepancies. Inter-rater reliability was calculated using the formula previously described [20]: Reliability = Agreement / (Agreement + Disagreement), yielding a reliability score of 90%, which exceeds the commonly accepted threshold of 70% in social research. Additionally, Cohen's Kappa coefficient was computed to provide a chance-corrected measure of agreement, with $\kappa = 0.88$, indicating a high level of coding consistency. In the quantitative phase, descriptive statistics were used to compare drawing-based data between participants diagnosed with schizophrenia and those without any psychiatric diagnosis. Group differences in categorical variables were examined using the Chi-Square Test, as this method is particularly suitable for assessing the significance of associations in qualitative data across independent groups [21].

RESULTS

This section presents findings based on the evaluation of data obtained from the "BND test" for the purpose of examining the attachment status of 50 participants diagnosed with schizophrenia and 50 participants without any psychiatric diagnosis (Table 1).

The evaluation of BND criteria revealed statistically significant differences in several core drawing features between individuals diagnosed with schizophrenia and those without such a diagnosis. Significant group differences were found for the presence of a parent bird in or near the nest ($\chi^2(1)= 4.52$, P=0,033, w= 0.21), the inclusion of a chick or egg ($\chi^2(1)= 4.54$, P=0,033, w= 0,21), green being the dominant color ($\chi^2(1)= 4.40$, P=0.036, w= 0.21), the appropriateness of coloring relative to the nature of the depicted objects ($\chi^2(1)= 21.72$, P<0.001, w= 0.47), the nest making contact with a visible surface ($\chi^2(1)= 22.04$, P<0.001, w= 0,47), and the depiction of both a parent bird and a chick/egg together ($\chi^2(1)= 11.63$, P<0.001, w= 0,34). In contrast, no significant differences were

Table 1. Comparison of attachment representations between participants with and without a schizophrenia diagnosis

BND Evaluation Criteria		Diagnosed with Schizophrenia	Not Diagnosed with Schizophrenia	χ2	P Value	Cohen's w
1. Is there a parent bird in or near the nest?	Yes	28 (56%)	39 (78)	4.52	**0.033***	0.213
	No	22 (44%)	11 (22%)			
2. Is there a chick or egg in or near the nest?	Yes	37 (74%)	46 (92%)	4.54	**0.033***	0.213
	No	13 (26%)	4 (8%)			
3. Is green the dominant color?	Yes	12 (24%)	23 (46%)	4.4	**0.036***	0.210
	No	38 (76%)	27 (54%)			
4. Is more than 20% of the paper surface used?	Yes	48 (96%)	42 (84%)	2.78	0.096	0.167
	No	2 (4%)	8 (16%)			
5. Is there a sun figure?	Yes	4 (8%)	9 (18%)	1.42	0.234	0.119
	No	46 (92%)	41 (82%)			
6. Is there a cloud figure?	Yes	4 (8%)	6 (12%)	0.11	0.738	0.033
	No	46 (92%)	44 (88%)			
7. Are birds drawn in an "M" shape?	Yes	3 (6%)	5 (10%)	0.14	0.712	0.037
	No	47 (94%)	45 (90%)			
8. Are there additional figures like flowers or pets?	Yes	18 (36%)	27 (54%)	2.59	0.108	0ş161
	No	32 (64%)	23 (46%)			
9. Is the coloring appropriate to the nature of the figures?	Yes	17 (34%)	41 (82%)	21.72	**0.001***	0.466
	No	33 (66%)	9 (18%)			
10. Is the nest enclosed within the trunk of a tree (if present)?	Yes	0 (0)	2 (4%)	0,51	0.475	0.071
	No	50 (100%)	48 (96%)			
11. Is the nest in contact with the ground (branch, leaf, etc.)?	Yes	18 (36%)	42 (84%)	22.04	**0.001***	0.469
	No	32 (64%)	8 (16%)			
12. Are parent and chick/egg birds drawn together?	Yes	18 (36%)	36 (72%)	11.63	**0.001***	0.341
	No	32 (64%)	14 (28)			

Data are shown as n (%). BND=Bird's Nest Drawing
*P<0.05.

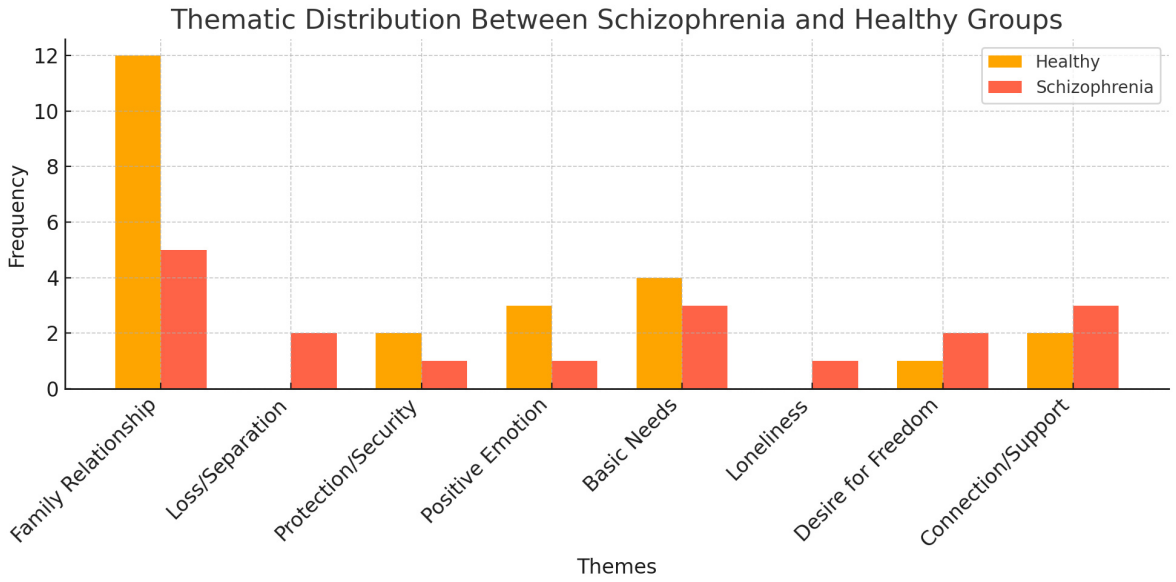

Fig. 1. Thematic distribution between the schizophrenia group and the control group.

observed between the groups in the use of more than 20% of the paper surface (P=0.096), inclusion of a sun figure (P=0.234), presence of clouds (P=0.738), "M"-shaped birds (P=0.712), additional figures such as flowers or pets (P=0.108), or nests enclosed within a tree trunk (P=0.475).

The quantitative findings indicate that significant differences exist between the two groups in terms of specific visual criteria related to the BND. However, in order to gain a deeper understanding of how these numerical differences reflect individuals' inner experiences, emotional patterns, and attachment representations, the written narratives accompanying the drawings were analyzed qualitatively. In this context, the narratives of 15 participants diagnosed with schizophrenia and 15 without a diagnosis were examined using discourse analysis. The aim was to explore how attachment representations are shaped through verbal and metaphorical expressions, and to identify themes that demonstrate similarities and differences between the two groups. The discourse themes, along with illustrative excerpts and content patterns from both groups, are presented below (Fig. 1).

The thematic analysis revealed that themes such as loneliness, desire for freedom, loss/separation, and basic needs were more frequently repeated in the stories of individuals diagnosed with schizophrenia. In contrast, themes such as family relationships, protection/security, and positive emotions were more prominent in the narratives of individuals who were not

Table 2. Group frequencies and descriptions by theme

Theme	Non-Schizophrenic group	Schizophrenic group
Family relationships	12	5
Loss/separation	0	2
Protection/security	2	1
Positive emotions	3	1
Basic needs	4	3
Loneliness	0	1
Desire for freedom	1	2
Connection/support	2	3

Fig. 2. **Sample bird's nest drawings by participants diagnosed with schizophrenia.**

diagnosed with schizophrenia.

When examining Table 2, themes such as access to basic needs, desire for freedom, loneliness, and loss are prominent in the stories of individuals diagnosed with schizophrenia. Metaphorical narratives such as "homelessness," "not being able to find food," "loneliness," and "spreading one's wings toward freedom," which are frequently encountered in these stories, indicate that the individual may have a weak secure attachment experience and emotional regulation difficulties. In addition, religious or mystical content has been observed in some narratives. In the written narratives of individuals who have not been diagnosed, the stories are generally shaped around family structure, caregiving, cohabitation, protection, and positive emotions. Motifs such as "the mother bird feeding her young," "family nest selection," and "spring and living in harmony with nature" provide clues about secure attachment representations and the capacity to form sustainable social relationships.

The final part of the findings section presents selected examples of bird's nest drawings created by both individuals with and without a schizophrenia diagnosis (Figs. 2 and 3).

DISCUSSION

This study examined the attachment representations of individuals diagnosed with schizophrenia and those without a diagnosis through the BND test and accompanying narratives. Quantitative findings revealed statistically significant differences between the two groups in several criteria, including the presence of parent and chick/egg bird figures, the use of green as a dominant color, appropriate coloring in line with the nature of the figures, and whether the nest was depicted in contact with a surface. Qualitative analysis showed that narratives by individuals with schizophrenia prominently featured themes such as loneliness, distrust, and loss. These findings indicate that individuals diagnosed with schizophrenia display more pronounced signs of insecure attachment representations.

In evaluating the drawings, the presence of parent birds, chick/egg figures, and similar components were notably less frequent among participants with schizophrenia compared to the non-clinical group, with these differences being statistically significant. These figures are often seen as indicators of familial unity and emotional closeness. Projective depictions of family-related elements can offer insights into individuals' internal representations of attachment and family dynamics. According to Akoğlu [22], in family-themed drawings, the presence, number, and spatial closeness of figures may reflect emotional bonds within the family.

Therefore, the significantly reduced inclusion of parent and chick/egg figures in drawings by partici-

Fig. 3. **Sample bird's nest drawings by participants without a schizophrenia diagnosis.**

pants with schizophrenia suggests possible deficits or weaknesses in their internalized family representations. Studies by Ebrinç *et al*. [23] and Karancı [24] emphasize that individuals diagnosed with schizophrenia frequently experience attachment-related traumas, emotional neglect, and inconsistent caregiving during childhood. Bowlby [2] argued that the absence of early secure attachment experiences can negatively impact the development of self-perception and trust in others. As a result, such individuals may perceive the external world as threatening and withdraw from relationships. The absence of parent and offspring bird figures in their drawings may thus be interpreted as a projective expression of these attachment-based difficulties. Furthermore, Tüzer *et al*. [25] highlight that families of individuals with schizophrenia often exhibit negative interaction patterns, including high expressed emotion, critical attitudes, and overprotectiveness, which may disrupt the individual's sense of safety and attachment system. Therefore, the lack of family figures in drawings should not be seen merely as an artistic limitation but rather as a clinically significant indication of restricted cognitive and emotional representations of familial experiences.

The findings also revealed significant differences in whether green was used as the dominant color and whether the coloring was appropriate to the nature of the figures. Green, symbolizing nature, vitality, and continuity of life, has been associated with secure attachment patterns in previous research [16, 17, 26]. In this study, participants with schizophrenia were found to use green less frequently and to demonstrate less consistency in coloring the figures appropriately compared to their non-diagnosed peers. This aligns with the findings of Shen *et al*. [27], who noted that individuals with schizophrenia tend to use fewer colors in projective drawings.

One of the most notable findings in this study was the frequent depiction of bird nests floating in space - disconnected from any visible ground - among individuals diagnosed with schizophrenia. When evaluated within the framework of Bowlby's [2] attachment theory, this situation shows that an individual's ability to position themselves on secure ground, emotionally and physically attached to a "support point," is based on early attachment experiences. Therefore, representations of a groundless home may reflect the individual's inability to construct a sense of inner security, to sufficiently internalize a sense of belonging, and to establish consistent bonds with their social environment. In this context, Liotti [6] states that early attachment traumas and unresolved experiences lead to fragmentation, inconsistency, and discontinuity in the individual's mental representations, deeply affecting their relationships with the outside world. Similarly, Uzun and Aslan [28] have shown in their pre-published studies that despite the desire of individuals diagnosed with schizophrenia to belong to society, experiences of exclusion and stigmatization seriously undermine this feeling. Barut *et al*. [29] also noted that themes of lack of belonging, loneliness, and social isolation are dominant in the lives of individuals on the schizophrenia spectrum. The fact that participants frequently describe themselves as "excluded," "lonely," and "not belonging anywhere" shows that the lack of belonging plays an important role in their mental structure. All these findings suggest that the presence of groundless nests in projective drawings is not merely an aesthetic choice but also a symbolic expression of the difficulties individuals experience in representing internal security, belonging, and relational integrity. In this context, the nests suspended in empty space appear not as mere artistic elements but as visual metaphors for emotional dislocation, relational fragmentation, and an enduring search for stability.

For some drawing criteria, no statistically significant differences were found between the groups. Specifically, the inclusion of elements such as the sun, clouds, "M"-shaped birds, flowers, or pets, the nest being enclosed in a tree trunk, and the proportion of paper used did not differentiate the groups significantly. These elements may reflect individual drawing style, personal expression, or aesthetic preference rather than attachment-related content. In projective assessments, the contextual and relational content of figures is more meaningful for psychodynamic interpretation, whereas elements that lack symbolic depth or direct relevance to the attachment system may have limited interpretability [30]. Therefore, the non-significant differences in these elements may be attributed to individual stylistic variation or non-attachment-based representations.

Qualitative findings further support the quantitative results, showing that individuals with schizophre-

nia expressed themes of unmet basic needs, desire for freedom, loneliness, and loss. Metaphorical expressions such as "lack of a nest," "inability to find food," "being alone," and "spreading wings toward freedom" suggest an unmet need for a secure attachment figure and underdeveloped emotional regulation mechanisms. According to attachment theory, a lack of trust-based early relationships with caregivers may lead to internal representations shaped by themes of distrust, abandonment, and isolation [31, 32]. In contrast, narratives from the non-clinical group included more positive themes such as "mother bird feeding her chicks," "building a nest as a family," and "living in harmony with spring and nature," reflecting secure attachment experiences, a sense of belonging, protection, and sustained social bonds. These contrasting themes suggest that individuals with schizophrenia exhibit signs of attachment disruptions and psychosocial deprivation, while the non-clinical group displays more coherent and positively internalized attachment systems.

Previous research employing various projective drawing techniques has also identified distinct visual and structural features associated with schizophrenia, supporting the present study's findings. For example, Kaneda *et al*. [33] compared tree-drawing test results between chronic schizophrenia patients and healthy controls, reporting significant differences in key structural indicators such as trunk width, base opening, branch-end size, and the ratio of tree area to paper area. These findings parallel the current study's observations that spatial organization and structural completeness in projective drawings may reflect underlying cognitive and emotional processing differences. Similarly, Teneycke *et al*. [34], using the bridge-drawing task, found that individuals with psychosis differed from control groups in formal art characteristics, including color choice, accuracy, and symbolic placement of elements such as the "future," suggesting that spatial-symbolic decisions in drawings may reveal altered perceptions and relational schemas. In a different cultural context, Iqbal *et al*. [35] analyzed emotional indicators in Draw-a-Person tests and observed that certain features - such as omission of key body parts, gross disproportions, and vacant eyes - occurred more frequently in the drawings of individuals with schizophrenia. Taken together, these studies align with the present findings by indicating that projective drawings, across diverse formats and cultural settings, capture meaningful markers of impaired integration, diminished symbolic coherence, and disrupted attachment representations in schizophrenia.

Limitations

While the findings contribute valuable insights to the literature, certain limitations should be acknowledged. First, the relatively small sample size and the recruitment of participants from a single institution limit the generalizability of the results. In particular, selecting participants exclusively from one private care center may have restricted the diversity of the sample in terms of socioeconomic background, cultural experiences, and clinical profiles. Second, the study relied solely on projective visual (BND) and narrative (storytelling) methods without incorporating standardized psychometric scales to measure attachment. As a result, the conclusions regarding attachment styles were based on indirect, interpretative analyses rather than objective, measurable data. Although projective techniques offer rich qualitative insights into the individual's inner world, they are limited in their ability to yield categorical classifications for complex constructs such as attachment. A more robust approach would have involved combining projective techniques with standardized attachment scales to enhance both the reliability and validity of the interpretations. Therefore, conclusions drawn regarding attachment representations should be interpreted cautiously and within the scope of these methodological limitations. Additionally, it should be noted that most participants with schizophrenia were on antipsychotic or other psychotropic medications, which may affect cognitive functioning, motor skills, and expressive behaviors (e.g., drawing performance, color use, narrative clarity). Thus, the performance exhibited in drawing and storytelling tasks may be influenced not only by psychopathology but also by medication effects. This presents another methodological limitation that complicates attributing projective findings solely to attachment representations.

Given the limitations noted in this study—such as single-site sampling, the absence of standardized attachment measures, and the potential influence of medication effects—future research should adopt mixed-method designs that integrate projective drawing and narrative techniques with validated attachment scales. Such an approach would not only strengthen

the validity and reliability of attachment assessment in schizophrenia but also bridge the gap between qualitative depth and quantitative precision. Longitudinal studies are particularly warranted to examine how attachment patterns influence clinical course, treatment adherence, relapse rates, and social reintegration over time. Future investigations should aim to control for neurobiological confounders, including the cognitive and expressive effects of psychotropic medication, by comparing medicated and non-medicated groups or statistically adjusting for medication status. Expanding recruitment to multiple sites with diverse cultural and socioeconomic backgrounds would improve the generalizability of findings and facilitate cross-cultural comparisons of attachment-related imagery. To ensure cross-cultural applicability, standardized measures should undergo cultural adaptation and validation prior to integration with projective techniques. From a methodological perspective, the symbolic and psychodynamic meanings of visual elements in projective drawings should be examined alongside thematic and contextual narrative analyses, potentially supported by computational tools such as visual content analysis software. Clinically, embedding attachment-sensitive principles into psychoeducation, individual therapy, and family interventions may foster more secure relational patterns and enhance psychosocial functioning. Finally, projective drawing techniques should be considered not only as diagnostic tools but also as therapeutic interventions that can promote emotional expression, strengthen self-awareness, and support the therapeutic alliance. Embedding these approaches into rehabilitation programs for individuals with schizophrenia may encourage insight, adaptive coping, and more holistic treatment outcomes.

CONCLUSION

In conclusion, the findings of this study suggest that the attachment representations of individuals with schizophrenia differ significantly from those without a diagnosis, both in visual drawings and narrative content. The presence or absence of specific figures in the BND, structural integrity of the drawings, and color choices were interpreted as projective reflections of the internal attachment system. The absence of parent and offspring bird figures, depiction of nests as floating and ungrounded, and limited use of appropriate coloring in the drawings of individuals with schizophrenia suggest weakened secure attachment experiences and vulnerabilities in emotional regulation and relational functioning. These findings were reinforced by qualitative data, which revealed themes of loneliness, rootlessness, unmet basic needs, and loss in the schizophrenia group. In contrast, drawings and narratives by the non-clinical group emphasized familial togetherness, caregiving, protection, and secure attachment patterns. Evaluated within the framework of attachment theory, all findings suggest that the qualitatively poor early caregiving experiences of individuals with schizophrenia significantly shape their internal representations, with implications for both psychopathology and social functioning.

Ethics Approval and Consent to Participate

This study was approved by the Istanbul Kent University Social and Human Sciences Ethics Committee (Decision No: 2025/06; date: 01.07.2025). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Written informed consent was obtained from all individual participants included in the study.

Data Availability

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

Authors' Contribution

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

Conflict of Interest

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

Financing

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

Acknowledgments

The author(s) have no acknowledgments to declare.

Generative Artificial Intelligence Statement

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

Editor's Note

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

REFERENCES

1. Ainsworth MDS, Blehar MC, Waters E, Wall S. Editors. Patterns of Attachment: A Psychological Study of the Strange Situation. Hillsdale, NJ: Lawrence Erlbaum, 1978.
2. Bowlby J. Attachment and Loss, Vol. 1: Attachment. Basic Books, 1969.
3. Cassidy J, Shaver PR. Handbook of Attachment: Theory, Research, and Clinical Applications. 2nd ed.. Guilford Press, 2008.
4. Mikulincer M, Shaver PR. Attachment in Adulthood: Structure, Dynamics, and Change. Guilford Press, 2007.
5. Dozier M, Stovall-McClough KC, Albus KE. Attachment and psychopathology in adulthood. In J. Cassidy & P. R. Shaver (Eds.), Handbook of Attachment. Guilford Press, 2008: pp. 718–744.
6. Liotti G. Trauma, dissociation, and disorganized attachment: three strands of a single braid. Psychotherapy. 2004;41(4):472-486. doi: 10.1037/0033-3204.41.4.472.
7. Berry K, Barrowclough C, Wearden A. A review of the role of adult attachment style in psychosis: unexplored issues and questions for further research. Clin Psychol Rev. 2007;27(4):458-475. doi: 10.1016/j.cpr.2006.09.006.
8. Mikulincer M, Shaver PR. An attachment perspective on psychopathology. World Psychiatry. 2012;11(1):11-15. doi: 10.1016/j.wpsyc.2012.01.003.
9. Karakuş G, Kocal Y, Sert D. [Schizophrenia: Etiology, Clinical Features and Treatment]. Arşiv Kaynak Tarama Dergisi. 2017;26(2): 251-267. doi: 10.17827/aktd.303574. [Article in Turkish]
10. Summakoğlu D, Ertuğrul B. [Schizophrenia and its treatment]. Lectio Scientific. 2018;2(1):43-61. [Article in Turkish]
11. Bifulco A, Kwon J, Jacobs C, Moran PM, Bunn A, Beer N. Adult attachment style as mediator between childhood neglect/abuse and adult depression and anxiety. Soc Psychiatry Psychiatr Epidemiol. 2006;41(10):796-805. doi: 10.1007/s00127-006-0101-z.
12. Read J, Gumley A. Can attachment theory help explain the relationship between childhood adversity and psychosis? In Telling Stories? Routledge, 2019: p. 51-94.
13. Nilsson A. Differentiation between patients with schizophrenia and borderline disorders in the Percept-genetic Object-Relation Test, PORT. Br J Med Psychol. 1995;68(4):287-309. doi: 10.1111/j.2044-8341.1995.tb01837.x.
14. Kaiser DH. Indications of attachment security in a drawing task. Arts Psychother.1996;23(4):333-340.
15. Demirbağ B. Kuş Yuvası Çizim Testinin Türkçeye Uyarlanması, Geçerlik ve Güvenirlik Çalışması. Yüksek Lisans Tezi, Ankara: Gazi Üniversitesi, 2016.
16. Akba Ö, Demir V. Analyzing Attachment Representations and Aggression Tendencies of Individuals Diagnosed with Psychosomatic Skin Disease through Drawing Method. Psikiyatride Güncel Yaklaşımlar. 2024;16(Suppl 1): 303-316. doi: 10.18863/pgy.1509412.
17. Gül K, Demir V. [Investigation of Attachment Representations and Personality Traits of Individuals with Alcohol and Substance Use Disorder by Drawing Pictures: Projective Research]. Bağımlılık Dergisi. 2025;26(1):47-58. doi: 10.51982/bagimli.1502012. [Article in Turkish]
18. Yıldırım A, Şimşek H. Sosyal bilimlerde nitel araştırma yöntemleri (11. baskı). Seçkin Yayıncılık, 2018.
19. Shenton AK. Strategies for ensuring trustworthiness in qualitative research projects. Educ Inform. 2004;22(2):63-75. doi: 10.3233/EFI-2004-22.
20. Miles MB, Huberman AM, Qualitative Data Analysis: An Expanded Sourcebook. 2nd ed. Calif.: SAGE Publications, 1994.
21. Dişçi R. Temel ve Klinik Biyoistatistik. İstanbul Medikal Yayıncılık, 2008.
22. Akoğlu B. Resim ve Çocuğun İç Dünyasına Yolculuk. İstanbul, Nobel Akademik Yayıncılık, 2023.
23. Ebrinç S, Çetin M, Başoğlu C, et al. [Research of family functionality, social support and expression of emotion in schizophrenic patients and their families]. Anadolu Psikiyatri Derg. 2001;2(1):5-14. [Article in Turkish]
24. Karancı AN. Şizofren hasta yakınlarının dünyası: Nedensel atıflar, yükler ve umutlar. Şizofreni Yazıları. 2000;1(1): 6-12.
25. Tüzer V, Zincir S, Başterzi AD, Aydemir Ç, Kısa C, Göka E. [Assessment of Expressed Emotion and Family Environment in Schizophrenic Patients]. Klinik Psikiyatri. 2003;6(4):189-197. [Article in Turkish]
26. Francis D, Kaiser D, Deaver SP. Representations of Attachment Security in the Bird's Nest Drawings of Clients with Substance Abuse Disorders. Art The. 2003;20(3):125-137. doi: 10.1080/07421656.2003.10129571.
27. Shen H, Wang SH, Zhang Y, et al. Color painting predicts clinical symptoms in chronic schizophrenia patients via deep learning. BMC Psychiatry. 2021;21(1):522. doi: 10.1186/s12888-

021-03452-3.
28. Uzun S, Aslan Ç. Schizophrenia Patients' Perceptions of Belongingness and Expectations from Society. Reserach Square. 2023. doi: 10.21203/rs.3.rs-2703461/v1.
29. Barut JK, Dietrich MS, Zanoni PA, Ridner SH. Sense of Belonging and Hope in the Lives of Persons with Schizophrenia. Arch Psychiatr Nurs. 2016;30(2):178-184. doi: 10.1016/j.apnu.2015.08.009.
30. Handler L, Thomas AD. Drawings in assessment and psychotherapy. New York: Routledge/Taylor & Francis Group, 2014.
31. Aydın AB. Yetişkin bireylerde erken dönem uyum bozucu şemalar ve bağlanma stilleri ile ayrılık kaygısı arasındaki ilişkinin incelenmesi. Yüksek lisans tezi. İstanbul Gelişim Üniversitesi, 2020.
32. Conradi HJ, Boertien SD, Cavus H, Verschuere B. Examining psychopathy from an attachment perspective: The role of fear of rejection and abandonment. J Forensic Psychiatry Psychol. 2016, 27(1): 92-109. doi: 10.1080/14789949.2015.1077264.
33. Kaneda A, Yasui-Furukori N, Saito M, et al. Characteristics of the tree-drawing test in chronic schizophrenia. Psychiatry Clin Neurosci. 2010;64(2):141-148. doi: 10.1111/j.1440-1819.2010.02071.x.
34. Teneycke T, Hoshino J, Sharpe D. The bridge drawing: An exploration of psychosis. Arts Psychother. 2009;36(5):297-303. doi: 10.1016/j.aip.2009.07.003.
35. Iqbal S, Shah AA, Iqbal M. Emotional indicators across Pakistani schizophrenic and normal individuals based on Draw a Person Test. Pak J Soc Clin Psychol. 2013;11(1): 59-65.

The European Research Journal 2025;11(5):1001-1007 Original Article

DOI: https://doi.org/10.18621/eurj.1754019 *Medical Oncology*

Is maintenance chemotherapy always necessary in gestational trophoblastic neoplasia? A retrospective cohort analysis

Emre Hafızoğlu[1], Murat Bardakçı[2], Efnan Algın[3], Öznur Bal[3], Doğan Uncu[3]

[1]Department of Medical Oncology, Afyonkarahisar State Hospital, Afyonkarahisar, Türkiye; [2]Department of Medical Oncology, Gazi Yaşargil Training and Research Hospital, Diyarbakır, Türkiye; [3]Department of Medical Oncology, Ankara Bilkent City Hospital, Ankara, Türkiye

ABSTRACT

Objectives: Gestational trophoblastic neoplasia (GTN) is a rare but highly curable group of gestational tumors. Current risk stratification relies on the International Federation of Gynecology and Obstetrics (FIGO) staging and WHO scoring systems, yet both have shown limited accuracy in predicting relapse or chemoresistance. The necessity of routine maintenance chemotherapy following remission - particularly in low-risk patients - remains controversial.

Methods: We conducted a retrospective cohort study of 25 patients with GTN treated between 2006 and 2022. Demographic, clinical, and treatment-related data were analyzed. Outcomes of interest included methotrexate (MTX) resistance, relapse, and the use of maintenance chemotherapy. Follow-up duration and disease outcomes were assessed descriptively.

Results: The median age at diagnosis was 28 years. Most patients (76%) had FIGO stage I disease; 44% were classified as high-risk. MTX resistance occurred in two patients (8%), both low-risk. Only one relapse was observed, occurring five years after remission. Maintenance chemotherapy was given to 64% of patients. Notably, none of the eight patients who did not receive maintenance therapy - including four high-risk cases - experienced relapse. No clear difference in outcomes was observed between stage I and stage III patients.

Conclusions: In this real-world cohort with long-term follow-up, maintenance chemotherapy did not appear necessary to prevent recurrence, even in select high-risk patients. Additionally, the FIGO/WHO systems showed limited prognostic discrimination. These findings support the need for individualized, response-adapted management strategies and underscore the limitations of current risk models in GTN.

Keywords: Maintenance chemotherapy, gestational trophoblastic neoplasia, FIGO stage

Gestational trophoblastic neoplasia (GTN) is a rare but highly curable group of gestational tumors, characterized by abnormal trophoblastic proliferation and elevated serum human chorionic gonadotropin (hCG) levels. With timely diagnosis and appropriate treatment, cure rates exceed 90%, even in metastatic disease [1]. Risk stratification typically relies on the International Federation of Gynecology and Obstetrics (FIGO) anatomical staging system and the World Health Organization (WHO) prognostic scoring

Received: June 30, 2025 **Accepted:** August 31, 2025 **Available Online:** September 1, 2025 **Published:** September 4, 2025

How to cite this article: Hafızoğlu E, Bardakçı M, Algın E, Bal Ö, Uncu D. Is maintenance chemotherapy always necessary in gestational trophoblastic neoplasia? A retrospective cohort analysis. Eur Res J. 2025;11(5):1001-1007. doi: 10.18621/eurj.1754019

Corresponding author: Emre Hafızoğlu, MD., Phone: +90 +90 444 42 34, E-mail: emrehafizoglu@gmail.com

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

system: low-risk patients (WHO score <7; FIGO stage I–III) are generally treated with single-agent methotrexate (MTX), while high-risk cases require multi-agent chemotherapy such as the etoposide-methotrexate-actinomycin D/cyclophosphamide-vincristine (EMA/CO) regimen [2, 3].

However, two aspects of GTN management remain controversial. First, the necessity of maintenance chemotherapy after remission is debated, especially in low-risk patients with inherently low relapse rates [4, 5]. Second, the prognostic accuracy of the FIGO/WHO system is increasingly questioned, with studies showing minimal differences in disease-free survival (DFS) across stages and relapse occurring even in low-risk patients [3-6].

Recent reviews have further highlighted the limitations of the FIGO 2000 system, citing methodological inconsistencies, ambiguous risk weighting, and failure to incorporate radiologic or molecular predictors. Jin-Kai *et al*. [6] proposed replacing this one-size-fits-all approach with task-specific prognostic models tailored to distinct clinical goals, such as predicting chemotherapy resistance or recurrence. These evolving perspectives support a shift toward more individualized, response-adapted strategies in GTN management.

In this study, we analyzed a 16-year cohort of GTN patients to evaluate relapse, while addressing two unresolved questions in GTN care - whether maintenance chemotherapy is essential for preventing recurrence, and whether FIGO staging accurately reflects clinical risk in long-term survivors.

METHODS

Study Design and Data Collection

This retrospective cohort study included all patients diagnosed with GTN between January 2006 and June 2022. Of 31 identified cases, 25 with complete data and adequate follow-up were included in the final analysis.

GTN was diagnosed based on clinical, biochemical, radiologic, and/or histopathologic findings. Included cases comprised postmolar GTN - defined by persistently elevated or rising β-hCG beyond six months after molar evacuation - and histologically confirmed choriocarcinoma, invasive mole, placental site trophoblastic tumor (PSTT), or epithelioid trophoblastic tumor (ETT). No patients with a WHO score >13 (ultra–high-risk category) were included in this study.

Demographic, clinical, and treatment-related variables were retrospectively collected, including patient characteristics, tumor pathology, treatment modalities, and outcomes such as MTX resistance, recurrence, and final disease status.

Low-risk patients were treated with single-agent MTX, while high-risk patients received multi-agent EMA/CO chemotherapy. Maintenance chemotherapy was defined as three additional cycles of the same regimen after complete remission, at the discretion of the treating physician.

Maintenance chemotherapy decisions were individualized, based on the depth and timing of hCG normalization, initial FIGO/WHO score, metastatic burden, treatment-related toxicity, patient compliance, and physician judgment. Patients with rapid and sustained hCG decline, minimal metastatic disease, and high toxicity risk were more likely to forgo maintenance therapy. Patients were monitored weekly with serum β-hCG until three consecutive values were below 5 mIU/mL, followed by monthly surveillance for at least six months. MTX resistance or progression was defined per FIGO 2002 criteria as a >10% rise over three values, plateau over four weeks, or persistent elevation beyond six months.

Risk Stratification and Staging Systems

Anatomical staging was performed according to the International Federation of Gynecology and Obstetrics (FIGO) 2002 criteria, which classifies gestational trophoblastic neoplasia (GTN) into four stages based on the anatomical extent of disease:

- **Stage I:** Disease confined to the uterus.
- **Stage II:** GTN extending to the genital structures (adnexa, vagina, or broad ligament).
- **Stage III:** Pulmonary metastases, with or without uterine involvement.
- **Stage IV:** Metastases to distant organs such as liver or brain.

Prognostic scoring was conducted using the FIGO/WHO 2000 risk scoring system, which integrates clinical, biochemical, and metastatic parameters

to predict resistance to single-agent chemotherapy. This composite score assigns numerical values (0, 1, 2, or 4 points) to the following eight prognostic (Table 1):

The cumulative score stratifies patients into risk groups as follows:

Low-risk: Total score 0-6

High-risk: Total score ≥7

Ultra-high-risk (optional classification): Score ≥13 or presence of extensive metastatic disease [2, 4, 5].

Accordingly, in our study, patients with FIGO stage I-III disease and a WHO score of 0-6 were classified as low-risk, whereas those with FIGO stage IV and/or a score ≥7 were classified as high-risk.

Ethical approval

All procedures involving human participants were conducted in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments. The study protocol was approved by the Ethics Committee of Ankara City Hospital (approval number: E1/2997/2022; date: 02/11/2022).

Statistical Analysis

All analyses were performed using IBM SPSS Statistics version 26.0 (IBM Corp., Armonk, NY). Continuous variables were reported as medians with ranges or interquartile ranges (IQRs), and categorical variables as counts and percentages. Primary outcomes included remission, recurrence, and MTX resistance. Follow-up duration was calculated from diagnosis to last clinical contact. Median follow-up and its 95% confidence interval (CI) were estimated using the non-parametric bootstrap method with 1,000 resamples. The observed follow-up range was also reported.

RESULTS

Twenty-five patients were included in the study, with a median age of 28 years. The most common presenting symptom was abnormal uterine bleeding, followed by asymptomatic β-hCG elevation. GTN developed after a molar pregnancy in slightly more than half of the cases. Most patients were diagnosed within four months of the antecedent pregnancy.

Invasive mole was the predominant histologic subtype. According to FIGO staging, most patients had stage I disease, and all metastases (n = 6) were limited to the lungs. Based on WHO/FIGO criteria, 56% of patients were classified as low-risk and 44% as high-risk (Table 2).

Chemotherapy was initiated in all but one patient, who underwent primary hysterectomy with subsequent remission. MTX was the first-line treatment in all low-risk and in selected high-risk patients. MTX resistance occurred in two cases, both successfully managed with multi-agent therapy. One relapse was documented five years after achieving initial remission.

Maintenance chemotherapy was administered in 64% of cases. Importantly, none of the eight patients

Table 1. Prognostic scoring

Risk Factor	Score 0	Score 1	Score 2	Score 4
Age (years)	<40	≥40	-	-
Antecedent pregnancy	Mole	Abortion	Term	-
Interval from index pregnancy (months)	<4	4-6	7-12	>12
Pretreatment serum hCG (IU/L)	$<10^3$	10^3–10^4	10^4–10^5	$>10^5$
Largest tumor size (including uterus) (cm)	<3	3-5	≥5	-
Site of metastases	Lung	Spleen, kidney	Gastrointestinal tract	Brain or liver
Number of metastases	0	1–4	5–8	>8
Previous failed chemotherapy	None	–	Single-agent	Multi-agent

Table 2. Baseline characteristics of patients with gestational trophoblastic neoplasia (n=25)

Characteristic	Data
Age at diagnosis (years)	28 (19-49)
Gravidity	3 (1-9)
Parity	1 (0-6)
Serum β-hCG at GTN diagnosis (IU/L)	8.5×10^5 (3.4×10^2 -1.2×10^7)
Presenting symptom	
Vaginal bleeding	12 (48.0%)
Asymptomatic (elevated hCG only)	9 (36.0%)
Amenorrhea	2 (8.0%)
Abdominal pain	2 (8.0%)
Antecedent pregnancy	
Molar pregnancy	13 (52.0%)
Complete mole	7 (28.0%)
Partial mole	6 (24.0%)
Non-molar pregnancy	12 (48.0%)
Abortion	3 (12.0%)
Ectopic pregnancy	2 (8.0%)
Term pregnancy	7 (28.0%)
Histologic subtype	
Invasive mole	23 (92.0%)
Choriocarcinoma	2 (8.0%)
FIGO anatomic stage	
Stage I	19 (76.0%)
Stage II	0 (0.0%)
Stage III	6 (24.0%)
Stage IV	0 (0.0%)
WHO/FIGO prognostic risk score	5 (2–10)
Low-risk (score <7)	14 (56.0%)
High-risk (score ≥7)	11 (44.0%)
Site of metastasis	
No metastasis	19 (76.0%)
Lung only	6 (24.0%)

Data are shown as n (%) or median (range)

who did not receive maintenance therapy - including four classified as high-risk - experienced relapse. This finding raises questions about the necessity of routine consolidation therapy in certain patients.

At a median follow-up of 30.7 months (95% CI: 27.7–55.5), all patients were alive and disease-free. No treatment-related deaths or severe adverse events were reported (Table 3).

No significant difference in long-term outcomes was observed between patients with FIGO stage I and III disease, suggesting limited prognostic discrimination in this cohort.

Table 3. Treatment characteristics and outcomes

Parameter	Data
First-line treatment	
Upfront hysterectomy	1 (4.0%)
Methotrexate	17 (68.0%)
EMA/CO regimen	7 (28.0%)
Maintenance chemotherapy	
Yes	16 (64.0%)
No	8 (32.0%)
Methotrexate resistance	2 (8.0%)
Recurrence	1 (4.0%)
Outcome	
Alive and disease-free	25 (100.0%)
Follow-up duration (months)	Median: 30.7 (95% CI: 27.7-55.5); range: 4.4-194.9

Data are shown as n (%). Follow-up duration is presented as median with 95% confidence interval and range.

Treatment characteristics and clinical outcomes of patients with gestational trophoblastic neoplasia (GTN). First-line therapies included single-agent methotrexate (MTX) in weekly or 3-day regimens, multi-agent EMA/CO, or primary surgery. Maintenance chemotherapy was defined as three additional cycles following remission. All patients were alive and disease-free at last follow-up.

DISCUSSION

This 16-year retrospective cohort provides real-world insight into two unresolved questions in GTN management: whether post-remission maintenance chemotherapy is essential and how accurately the FIGO/WHO risk stratification systems predict long-term outcomes.

Several studies have questioned the discriminatory power of the FIGO and WHO scoring systems in guiding therapeutic decisions. Powles *et al*., in a large retrospective series of 1708 cases, identified 60 relapses and 11 chemoresistant cases, noting that relapses occurred even among low-risk patients and that prognostic scores poorly predicted survival after relapse, especially in ultra–high-risk disease [6]. Similarly, Osborne *et al*. [7] evaluated 216 low-risk GTN patients enrolled in a randomized trial and observed methotrexate resistance in patients with WHO scores of 5–6, suggesting inadequate discrimination within the low-risk group. A meta-analysis including 901 complete responders by Albright *et al*. [8] reported a 4.1% relapse rate, with markedly higher mortality in patients with ultra–high-risk scores (≥13), despite comparable treatment regimens [8]. In their comprehensive guideline, Seckl *et al*. [9] also noted the limited prognostic performance of anatomical staging alone and recommended incorporating clinical judgment and hCG kinetics.

Our own findings further support these concerns. No significant difference in long-term outcomes was observed between patients with FIGO stage I and III disease, and both instances of methotrexate resistance occurred in patients classified as low-risk. These results challenge the reliability of current staging and scoring systems in capturing the true biological heterogeneity of GTN.

Importantly, recent work by Jin-Kai *et al*. [6] offers a deeper critique of the FIGO 2000 framework. The authors argue that the system was originally developed to describe general prognosis, not to inform clinical decisions such as predicting resistance or recurrence. They also demonstrate that the WHO score relies on arbitrarily assigned weighting and lacks statistical grounding, which may lead to risk misclassification. Their findings highlight the need for task-specific, biologically informed prognostic models tailored to discrete clinical endpoints [6]. This aligns closely with our observations and reinforces the need for refined stratification tools that integrate clinical, biochemical, and potentially molecular variables.

The second area of controversy—maintenance chemotherapy—remains unsettled. Expert consensus often favors continued treatment for several weeks post-remission, particularly in high-risk patients [9]. However, this practice is based largely on retrospective data rather than prospective validation. Albright *et al*. [8] reviewed over 2,100 high-risk patients and found limited evidence supporting uniform consolidation, despite its widespread adoption. Our study contributes to this debate by showing that none of the eight patients who did not receive maintenance therapy—including four high-risk cases—developed relapse during long-term follow-up.

This finding is supported by several reports questioning the need for three-cycle consolidation protocols. In the MITO-9 study, a multicenter retrospective analysis of 333 low-risk GTN patients treated with first-line methotrexate across six institutions in Italy. Following hCG normalization, patients were grouped according to the number of consolidation cycles received—two, three, or more than three—and relapse rates were assessed. Notably, no relapses occurred among patients with FIGO scores ≤2 who received only two consolidation cycles. In contrast, relapse rates increased to 2.2% and 10.2% among those receiving three and more than three cycles, respectively, regardless of FIGO subscore [10]. Conversely, Mitric *et al*. [11] conducted a retrospective analysis of 94 patients across two Canadian academic centers and demonstrated excellent outcomes with standardized three-cycle consolidation following care centralization in Canada. Several institutions, including Braga *et al*. [12], continue to advocate for three consolidation cycles following remission in both low- and high-risk GTN patients. Lybol *et al*. [13] also reported lower relapse with three versus two methotrexate cycles (4.0% vs. 8.3%, P=0.006). However, Couder *et al*. [14], in a large cohort of 465 patients, found that the number of consolidation cycles was not independently predictive of recurrence; rather, antecedent term pregnancy and the need for ≥5 MTX cycles were stronger indicators.

Taken together, these findings - including those from our cohort - underscore the need to move beyond rigid, uniform maintenance protocols. Personalized, response-adapted strategies based on dynamic treatment indicators may provide safer and more effective post-remission care. In line with this, Jin-Kai *et al*. [6] proposed integrating innovative tools such as slope modeling, time-series analysis, and machine learning into future prognostic models. Similarly, a recent systematic review protocol highlighted the absence of high-quality evidence supporting universal consolidation and advocated for individualized management in postmolar GTN [15].

Limitations

This study has several limitations. Its retrospective design limits causal inference and introduces potential selection bias. The small sample—especially in subgroup analyses—diminishes statistical power. Additionally, being a single-center study may limit generalizability. Nonetheless, the findings provide meaningful real-world insights and raise clinically relevant questions for future prospective research.

CONCLUSION

Our real-world data support two central insights in GTN management. First, maintenance chemotherapy may be safely omitted in selected—including some classified as high-risk—without compromising long-term outcomes. Second, the FIGO/WHO scoring systems showed limited prognostic accuracy in predicting relapse or resistance. These findings emphasize the need for individualized, response-adapted risk models that integrate clinical dynamics and data-driven predictors. Prospective multicenter studies are warranted to validate these strategies and guide optimal post-remission care.

Ethical Statement

This study was approved by the Ankara City Hospital Clinical Research Ethics Committee No. 1 (decision no.: E1/2997/2022; approval date: 02.11.2022). The study was conducted in accordance with the principles of the Declaration of Helsinki and with good clinical practice guidelines. The requirement for informed consent was waived due to the retrospective design of the study and the use of anonymized clinical data.

Data Availability

All data generated or analyzed during this study are included in this published article. Additional datasets that support the findings of this study are

available from the corresponding author upon reasonable request.

Authors' Contribution

Study Conception: EA, DU; Study Design: EA, DU; Supervision: DU; Funding: MB, ÖB, EA; Materials: EH; Data Collection and/or Processing: MB, EH; Statistical Analysis and/or Data Interpretation: ÖB, EA; Literature Review: ÖB, EA; Manuscript Preparation: EH; and Critical Review: EH.

Conflict of interest

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

Financing

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

Acknowledgments

The authors have no acknowledgments to declare.

Generative Artificial Intelligence Statement

Portions of the manuscript were developed with the assistance of ChatGPT-4o (May 2024 version), a large language model developed by OpenAI. This tool was employed solely to enhance linguistic clarity, improve academic expression, and ensure stylistic consistency. All medical content, statistical analyses, and scientific interpretations were independently generated by the authors. The final version of the manuscript was thoroughly reviewed and approved by all authors.

Editor's note

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

REFERENCES

1. Lurain JR. Gestational trophoblastic disease I: epidemiology, pathology, clinical presentation and diagnosis of gestational trophoblastic disease, and management of hydatidiform mole. Am J Obstet Gynecol. 2010;203(6):531-539. doi: 10.1016/j.ajog.2010.06.073.

2. FIGO Oncology Committee. FIGO staging for gestational trophoblastic neoplasia 2000. FIGO Oncology Committee. Int J Gynaecol Obstet. 2002;77(3):285-287. doi: 10.1016/s0020-7292(02)00063-2.

3. Niemann I, Vejerslev LO, Frøding L, et al. Gestational trophoblastic diseases - clinical guidelines for diagnosis, treatment, follow-up, and counselling. Dan Med J. 2015;62(11):A5082.

4. Ngan HYS, Seckl MJ, Berkowitz RS, et al. Update on the diagnosis and management of gestational trophoblastic disease. Int J Gynaecol Obstet. 2018;143 Suppl 2:79-85. doi: 10.1002/ijgo.12615.

5. Powles T, Savage PM, Stebbing J, et al. A comparison of patients with relapsed and chemo-refractory gestational trophoblastic neoplasia. Br J Cancer. 2007;96(5):732-737. doi: 10.1038/sj.bjc.6603608.

6. Jin-Kai L, Fang J, Yang X. Prognosticating gestational trophoblastic neoplasia: from FIGO 2000 to future models. EClinicalMedicine. 2024;77:102890. doi: 10.1016/j.eclinm.2024.102890.

7. Osborne RJ, Filiaci V, Schink JC, et al. Phase III trial of weekly methotrexate or pulsed dactinomycin for low-risk gestational trophoblastic neoplasia: a gynecologic oncology group study. J Clin Oncol. 2011;29(7):825-831. doi: 10.1200/JCO.2010.30.4386.

8. Albright BB, Ellett T, Knochenhauer HE, et al. Treatments and outcomes in high-risk gestational trophoblastic neoplasia: A systematic review and meta-analysis. BJOG. 2023;130(5):443-453. doi: 10.1111/1471-0528.17374.

9. Seckl MJ, Sebire NJ, Berkowitz RS. Gestational trophoblastic disease. Lancet. 2010;376(9742):717-729. doi: 10.1016/S0140-6736(10)60280-2.

10. Cioffi R, Fruscio R, Sabetta G, et al. Consolidation courses in low-risk gestational trophoblastic neoplasia and relapse rate: A MITO-9 retrospective study. Gynecol Oncol. 2025;196:54-58. doi: 10.1016/j.ygyno.2025.03.033.

11. Mitric C, Yang K, Bhat G, et al. Gestational trophoblastic neoplasia: does centralization of care impact clinical management? Int J Gynecol Cancer. 2023;33(11):1724-1732. doi: 10.1136/ijgc-2023-004526.

12. Braga A, Elias KM, Horowitz NS, Berkowitz RS. Treatment of high-risk gestational trophoblastic neoplasia and chemoresistance/relapsed disease. Best Pract Res Clin Obstet Gynaecol. 2021;74:81-96. doi: 10.1016/j.bpobgyn.2021.01.005.

13. Lybol C, Sweep FC, Harvey R, et al. Relapse rates after two versus three consolidation courses of methotrexate in the treatment of low-risk gestational trophoblastic neoplasia. Gynecol Oncol. 2012;125(3):576-579. doi: 10.1016/j.ygyno.2012.03.003.

14. Couder F, Massardier J, You B, et al. Predictive factors of relapse in low-risk gestational trophoblastic neoplasia patients successfully treated with methotrexate alone. Am J Obstet Gynecol. 2016;215(1):80.e1-7. doi: 10.1016/j.ajog.2016.01.183.

15. Branco-Silva M, Maesta I, Elias K, Berkowitz RS, Abbade JF, Horowitz NS. Consolidation chemotherapy in postmolar low-risk gestational trophoblastic neoplasia: a systematic review protocol. BMJ Open. 2022;12(2):e059484. doi: 10.1136/bmjopen-2021-059484.

The European Research Journal 2025;11(5):1008-1014 Original Article

DOI: https://doi.org/10.18621/eurj.1750403 *Gastroenterological Surgery*

Comparison of clinical outcomes of Lichtenstein and laparoscopic totally extraperitoneal repair techniques in inguinal hernia surgery

Ramazan Saygın Kerimoğlu[1], Süleyman Kargın[2], Abdullah Sami Maden[3], Abdullah Gürhan Duyan[3]

[1]Department of Gastroenterological Surgery, Division of Bariatric and Metabolic Surgery, Konya City Hospital, Konya, Türkiye; [2]Department of General Surgery, Division of Bariatric and Metabolic Surgery, Konya City Hospital, University of Health Sciences, Konya, Türkiye; [3]Department of General Surgery, Konya City Hospital, Konya, Türkiye

ABSTRACT

Objectives: This study aims to compare the clinical outcomes of open surgery (Lichtenstein procedure) with laparoscopic surgery (Totally Extraperitoneal Repair - TEP) for inguinal hernia repair. The duration of the procedure, postoperative pain, and occurrence of complications were evaluated.

Methods: A retrospective research including 95 male patients was carried out from June 2023 to January 2025. Patients were classified into Lichtenstein (n=50) and TEP (n=45) groups. Demographic data, operational durations, postoperative pain evaluations (visual analog scale - VAS), return-to-work intervals, and complications were analyzed for comparison.

Results: The operating duration was diminished in the TEP group, with 24±3.1 minutes for unilateral patients and 37±4.2 minutes for bilateral cases, in contrast to 46±5.3 minutes and 72±8.0 minutes in the Lichtenstein group, respectively (P<0.001). Postoperative pain scores were significantly decreased in the TEP group at all time intervals (P<0.001). The recurrence rates were low in both cohorts, recorded at 2.0% for the Lichtenstein group and 2.2% for the TEP group. The complication rates were similar, with wound infection occurring in 6% and hematoma in 4% of the Lichtenstein group, whereas seroma was noted in 4.4% of the TEP group.

Conclusions: The laparoscopic TEP technique surpasses open surgery in terms of operating length and postoperative pain; yet, both methods are safe therapeutic options with low complication and recurrence rates.

Keywords: Inguinal hernia, Lichtenstein procedure, postoperative pain, totally extraperitoneal repair

Inguinal hernia is a common concern in surgical practice, with approximately 20 million operations performed worldwide each year [1, 2]. The lifetime prevalence is reported to be 27% in males and 3% in females. Tension-free techniques, notably the Lichtenstein Procedure (LP), an open surgical approach established in the 1980s, have been regarded as the gold standard in hernia repair owing to its minimal recurrence rates and straightforward implementation [1, 4]. However, open surgical methods include disadvantages, such as chronic pain and prolonged recovery durations [5].

Received: July 27, 2025 **Accepted:** August 31, 2025 **Available Online:** September 1, 2025 **Published:** September 4, 2025

How to cite this article: Kerimoğlu RS, Kargın S, Maden AS, Duyan AG. Comparison of clinical outcomes of Lichtenstein and laparoscopic totally extraperitoneal repair techniques in inguinal hernia surgery. Eur Res J. 2025;11(5):1008-1014. doi: 10.18621/eurj.1750403

Corresponding author: Abdullah Gürhan Duyan, MD., Phone: +90 332 310 50 00, E-mail: dr.gurhanduyan@gmail.com

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

Laparoscopic techniques, including Totally Extraperitoneal Repair (TEP) and Transabdominal Preperitoneal Repair (TAPP), have arisen as alternatives to open surgery. Laparoscopic treatments offer advantages in postoperative pain management, cosmetic results, and the time required to return to normal activities [3]. Laparoscopic procedures, however, have disadvantages, such as a prolonged learning curve and potentially longer operation durations compared to open techniques [2]. Moreover, data suggest that laparoscopic repairs demonstrate reduced recurrence rates compared to open surgery when performed by proficient surgeons [6].

Nonetheless, it is underscored that each approach possesses distinct advantages and disadvantages across various patient populations. This study is to evaluate the results of open versus laparoscopic surgical procedures. The study evaluates the effectiveness of both operations in relation to key criteria such as operational length, postoperative pain, and recurrence rates. This study aims to provide new data to guide the choice of surgical procedures based on current literature.

METHODS

This retrospective study sought to evaluate 95 inguinal hernia surgeries performed between June 2023 and January 2025. The study compared open surgery using the LP technique with laparoscopic surgery with the TEP method. The research received approval from the hospital ethics committee (document number: 2025/023, date: 12 June 2025). The research was performed in compliance with the 1964 Helsinki Declaration and its subsequent revisions or comparable ethical norms.

Criteria for Patient Selection and Study Cohort

Inclusion criteria included the following: (1) Male patients diagnosed with inguinal hernia, (2) Aged between 18 and 72, (3) Eligible for elective surgery, and (4) Hernia size <3 centimeter. Exclusion criteria also was (a) Female patients, (b) Surgeries employing the Zig technique, (c) Complicated hernias (incarcerated, strangulated, recurrent), (d) Scrotal hernia, (e) Femoral Hernias, (f) Patients with lung disease, diabetes, and connective tissue diseases, (g) Smoking patients, and (h) Patients who wish to discontinue follow-up.

Study population:

A total of 95 patients were operated on in both groups: 45 patients in the laparoscopic TEP group, and 50 patients in the LP group.

Surgical Techniques

The procedure involves a 14×9 cm polypropylene mesh placed over the Poupart ligament and conjoined tendon, secured with 2-0 Prolene, followed by a typical tension-free mesh repair. The procedures were performed under spinal anesthesia.

In the TEP procedure, a 10 mm blunt trocar was inserted following a 1 cm infraumbilical incision and blunt dissection of the rectus muscles to expose the posterior rectus sheath. Two 5-mm working ports were inserted in the midline. Blunt dissection of the preperitoneal cavity was performed. Following thorough dissection down Cooper's ligament to the femoral canal, the hernia sac was fully dissected and entirely reduced. The TEP technique entailed the insertion of a 14×9 cm mesh in the preperitoneal space, affixed with two absorbable tacks across the Cooper ligament under general anesthesia. The insufflation pressure was sustained at 12 mmHg.

Postoperatively, 500 milligrams (mg) of paracetamol was provided intravenously every 8 hours, along with 50 mg dexketoprofen every 12 hours. Postsurgery, patients received slip underwear one size smaller than their usual fit and were instructed to wear them for 10 days to mitigate scrotal edema. Furthermore, 3-kilogram weights were applied to the groin regions of the patients for the initial 4 hours postsurgery. Following a duration of 4 hours, patients were mobilized, and oral intake commenced. All patients were released the day following the procedure. Upon discharge, a prescription for 50 mg of oral dexketoprofen was issued.

Assessment Criteria

1. Duration of operation

2. Postoperative pain (VAS= Visual Analog Scale)

3. Incidence of complications (seroma, hematoma, infection, etc.)

4. Rates of recurrence

5. Time to return work/day (The patients in each group were divided into two groups according to job types. Job type 1 included desk workers, and job type 2 included heavy manual workers)

The Visual Analog Scale (VAS)

VAS is a subjective measurement instrument frequently utilized to determine postoperative pain levels, and its reliability as a pain assessment tool has been substantiated in clinical studies [7]. This scale necessitates that patients evaluate their pain on a continuum from 0 to 10, where "0" signifies the absence of pain and "10" indicates the most severe pain. It is simple and readily adaptable nature makes it widely employed in both surgical clinics and research settings [8]. In our study, we assessed the scores at 6, and 24 hours post-surgery while the patients remained hospitalized. Upon reevaluation during outpatient visits on the 10th day, we scheduled follow-up sessions for the patients 6 months later. Patients who failed to attend their 6-month follow-up appointment were contacted by telephone to get a VAS score. Unreachable patients were excluded from the research.

Statistical Analysis

The data were analyzed using SPSS version 25.0. Continuous data are presented in the form of mean± standard deviation and compared using an independent t test, whereas categorical data are presented in frequency (%) and compared using the chi-square test. Independent sample t-tests and Chi-square testing were employed, with $P<0.05$ considered statistically significant.

RESULTS

The average age of 95 patients was 38.4 years, ranging from 18 to 72 years. No significant difference was observed between the groups regarding age and body mass index (BMI) ($P>0.05$).

The TEP method has attracted interest because to its shorter operational duration compared to the LP. The literature suggests that open repairs are typically shorter. The mean length in the TEP group was 24±3.1 minutes for unilateral instances and 37±4.2 minutes for bilateral occurrences. In contrast, the durations in the Lichtenstein group were 46±5.3 minutes and 72±8.0 minutes, respectively. The difference between the two methods was statistically significant ($P<0.001$). The results demonstrate that the TEP approach surpasses open surgery in terms of operating speed (Table 1).

Postoperative pain was evaluated by the Visual

Table 1. Demographic details of patients according to groups

Variables	LP group (n=50)	TEP group (n=45)	P value
Age (years)	39.2±11.3	37.5±10.8	0.421
BMI (kg/m²)	24.7 3.2	24.9±2.8	0.552
Location			
Right side	23 (46%)	14 (31.1%)	0.341
Left side	7 (14%)	11 (24.4%)	0.712
Bilateral	20 (40%)	20 (44.5%)	0.975
Type			
Indirect hernia	38 (76%)	32 (71%)	0.162
Direct hernia	12 (24%)	13 (29%)	0.491
Duration			
Unilateral (min)	46±5.3	24± 3.1	**0.001**
Bilateral (min)	72±8.0	37±4.2	**0.001**

Data are shown as mean±standard deviation or n (%). BMI=Body mass index, LP=Lichtenstein Procedure, TEP=Totally Extraperitoneal Repair

Table 2. VAS pain scores in postoperative follow-up according to groups

Postoperative time (min)	LP group	TEP group	P value
6th hours	6.9±1.2	4.1±0.9	**<0.001**
24th hours	5.7±1.1	3.3±0.8	**<0.001**
10th days	2.9±0.7	1.4±0.5	**<0.001**
6th months	1.5±0.4	0.8±0.3	**<0.001**

Data are shown as mean±standard deviation. LP=Lichtenstein Procedure, TEP=Totally Extraperitoneal Repair, VAS=Visual Analog Scale

Analog Scale after 6 hours, 24 hours, 10 days, and 6 months post-surgery. Pain scores were heightened in the LP at all time intervals. The TEP group exhibited a significantly decreased level of pain relative to open surgery across all intervals (Table 2).

In the LP group, wound infection was observed in 3 (6%) patients and hematoma in 2 (4%) patients. Seroma (4.4%) was identified in two participants in the TEP cohort. No significant difference was noted in complication rates (P>0.05).

During the 6-month follow-up period, recurrence was observed in 1 (2.0%) patient in the LP group and in 1 (2.2%) patient in the TEP group. Recurrence was noted at 2 months postoperatively for the LP method and at 1 month for the TEP procedure. No statistically significant difference in recurrence rates was observed between the two methods (P=0.67) (Table 3). The results demonstrate that both surgical methods offer a reliable therapeutic option with low recurrence rates.

Return-to-work times are provided in Table 4. The average time to return to work was considerably shorter in the TEP group (12.8±7.1 days) compared to the open group (19.3±4.3; P<0.001). The duration until return to work was significantly shorter for patients in both occupational classifications within the TEP group.

DISCUSSION

While the Lichtenstein procedure remains the gold standard for inguinal hernia surgery, the favorable outcomes associated with laparoscopic techniques have prompted renewed concerns regarding the optimal surgical approach. The objective of our study was to enhance these conversations by contrasting the outcomes of the laparoscopic technique with those of the gold standard LP. These data demonstrate the advantages of laparoscopic surgery and suggest that both procedures are relevant to appropriate patient populations. Our data demonstrate that the TEP technique offers significant advantages for operational length, postoperative discomfort, and time to return to work.

Assessments of recurrence rates demonstrated diversity in both magnitude and direction of effects. The majority of assessments of primary hernias, recurrent hernias, and both categories combined were rated as high to moderate quality, signifying considerable confidence in their conclusions. Supplementary reviews lacking specification of hernia types produced equivocal results, with four reviews [9-12] of moderate to low quality indicating statistically significant outcomes favoring open repair approaches; still, the overall low quality of these reviews raises considerable concern.

Table 3. Postoperative complications

Complications	LP group (n=50)	TEP group (n=45)	P value
Wound infections	3 (6%)	0	0.18
Hematoma	2 (4%)	0	0.32
Seroma	0	2 (4.4%)	0.27
Recurrences	1 (2%)	1 (2.2%)	0.67

Data are shown as n (%). LP=Lichtenstein Procedure, TEP=Totally Extraperitoneal Repair

Table 4. Return of work (days) details according to groups

LP group			TEP group		
Job type 1 (n=28)	**Job type 2 (n=22)**	**Total Job (n=50)**	**Job type 1 (n=31)**	**Job type 2 (n=14)**	**Total Job (n=45)**
14.25±4.2 (7-16)	25.52±9.6 (7-40)	19.3±4.3 (7-40)	9.27±3.0 (5-13)	17.57±6.7 (12-30)	12.8±7.1 (5-30)

Data are shown as mean±standard deviation (range). LP=Lichtenstein Procedure, TEP=Totally Extraperitoneal Repair

The results correspond with those outlined in a recent clinical guideline [13], in which experienced hernia surgeons and researchers worldwide offered recommendations based on comparable recurrence rates between laparoscopic and open surgical procedures. The recommendation emphasized that recurrence rates remain comparable, especially when procedures are conducted by highly proficient surgeons.

A meta-analysis shown that laparoscopic repairs are consistently associated with decreased pain relative to open repairs for chronic groin pain, regardless of the repair technique or hernia type examined [14]. A recent analysis defined both early and late chronic groin pain, however the conclusions are consistent with those of other publications in this overview [15]. Two reviews were seen to include identical research and individuals across many comparisons of hernia types, whereas their conclusions are consistent with those of other reviews [16, 17].

In a study conducted by Ulusoy *et al*. [18], TEP surgery was found to be superior than LP surgery in the postoperative period, corroborating our findings [18]. Numerous contemporary publications have demonstrated that surgical durations are comparable or even reduced, with complication rates occurring at similar frequencies. In their investigation on geriatric patients, Hernandez-Rosa *et al*. [19] and Çiftçi [20] and noted that the durations of open and laparoscopic surgeries were comparable [19, 20]. Our analysis indicated reduced surgical durations and comparable complication rates, aligning with contemporary literature. Reduced surgery durations were likewise noted in bilateral surgery cohorts. Our study links this variance to the surgeon's proficiency in laparoscopic methods. Lower VAS scores in the TEP group across all follow-up periods indicated the beneficial impact of laparoscopic techniques on acute and chronic pain problems. This scenario frequently signifies that the learning curve for laparoscopic techniques has been surmounted, resulting in a notable impact on surgical durations and complication rates.

A significant benefit of laparoscopic techniques is their facilitation of a prompt return to employment. Conflicting articles exist on this subject. Shah *et al*. [5] showed in a prospective randomized research that the return to work following LP was shorter. Numerous studies have demonstrated that laparoscopic techniques facilitate an expedited return to work [21]. This study further demonstrates that laparoscopic techniques facilitate a more expedited return to employment.

Limitations

A major limitation of our study is the lack of an extended follow-up period necessary to detect recurrent occurrences. However, since the study primarily focused on surgical complications and postoperative comfort, the follow-up duration was shortened. The investigation of recurrence in these patients may necessitate a separate study. A notable disadvantage is that patients in the open group underwent surgery with spinal anesthesia, while those in the laparoscopic group received general anesthesia. The variations in anesthetic techniques among the groups may have impacted the evaluated measures. Investigating the differences between these surgical approaches under uniform anesthesia protocols should yield more reliable results.

CONCLUSION

This study compared the efficacy of open surgery vs laparoscopic surgery in treating inguinal hernia. The

findings indicated that laparoscopic surgery excelled in operative duration, return to work duration, and postoperative discomfort. The complication rates for both treatments were minimal, and the recurrence rate was low. These results contribute to the good outcomes of laparoscopic methods and support laparoscopic methods in the debate over the gold standard treatment method. Subsequent research using bigger patient cohorts and extended follow-up may enhance the validity of these results. Additionally, it is advisable to do a cost-effectiveness analysis and further examine the impact on quality of life.

Ethics Approval and Consent to Participate

This study was approved by the KTO Karatay University, School of Medicine, Non-Interventional Clinical Research Ethics Committee (Decision No: 2025/023; Date: 12.06.2025). All procedures were conducted in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration and its later amendments. Informed consent was waived because of the retrospective nature of the study. All data were anonymized, and participant confidentiality was strictly maintained.

Data Availability

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

Authors' Contribution

Study Conception: RSK, SK, AGD; Study Design: RSK, SK, AGD; Supervision: RSK, SK; Funding: N/A; Materials: N/A; Data Collection and/or Processing: SK, ASM, AGD; Statistical Analysis and/or Data Interpretation: SK, ASM, AGD; Literature Review: RSK, SK, ASM, AGD; Manuscript Preparation: SK, ASM, AGD and Critical Review: RSK, SK.

Conflict of Interest

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

Financing

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

Acknowledgments

The author(s) have no acknowledgments to declare.

Generative Artificial Intelligence Statement

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

Editor's Note

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

REFERENCES

1. Vallipriya V, Surendran S. Comparison of outcome between laparoscopic inguinal hernia repair and open inguinal hernia repair. Int Surg J. 2022;9(6):1231-1235. doi: 10.18203/2349-2902.isj20221335.
2. Burton V, Perez AJ. Comparison of open and laparoscopic inguinal hernia repair. Mini-invasive Surg. 2021;5:26. doi: 10.20517/2574-1225.2021.26.
3. Pulikkal Reghunandanan R, Ali Usman A, Basheer S, Kuttichi L, Els Jojo J, Abdul Rasheed MF. Laparoscopic Versus Open Inguinal Hernia Repair: A Comparative Study. Cureus, 2023;15(11):e48619. doi: 10.7759/cureus.48619.
4. Bind RK, Ram B, Singh U. Direct Inguinal Hernias and Anterior Surgical Approach are risk factors for female Inguinal Hernia Recurrences. Int J Life Sci Biotechnol Pharma Res. 2024;13(9):97-101. doi: 10.69605/ijlbpr_13.9.2024.16.
5. Shah MY, Raut P, Wilkinson TRV, Agrawal V. Surgical outcomes of laparoscopic total extraperitoneal (TEP) inguinal hernia repair compared with Lichtenstein tension-free open mesh inguinal hernia repair: A prospective randomized study. Medicine (Baltimore). 2022;101(26):e29746. doi: 10.1097/MD.0000000000029746.
6. Zhao Y, Xu Z, Wang T, et al. The impact of laparoscopic versus open inguinal hernia repair for inguinal hernia treatment: A retrospective cohort study. Health Sci Rep. 2023;6(4):e1194. doi: 10.1002/hsr2.1194.
7. Hawker GA, Mian S, Kendzerska T, French M. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). Arthritis Care Res (Hoboken).

2011;63 Suppl 11:S240-52. doi: 10.1002/acr.20543.
8. Collins SL, Moore RA, McQuay HJ. The visual analogue pain intensity scale: what is moderate pain in millimetres? Pain, 1997;72(1-2):95-97. doi: 10.1016/s0304-3959(97)00005-5.
9. Memon MA, Cooper NJ, Memon B, Memon MI, Abrams KR. Meta-analysis of randomized clinical trials comparing open and laparoscopic inguinal hernia repair. Br J Surg. 2003;90(12):1479-1492. doi: 10.1002/bjs.4301.
10. Gavriilidis P, Davies RJ, Wheeler J, de'Angelis N, Di Saverio S. Total extraperitoneal endoscopic hernioplasty (TEP) versus Lichtenstein hernioplasty: a systematic review by updated traditional and cumulative meta-analysis of randomised-controlled trials. Hernia 2019;23(9):1093-1103. doi: 10.1007/s10029-019-02049-w.
11. O A, A G, M J, H WC, Al-Kandari A, S C, M M. Is laparoscopic inguinal hernia repair more effective than open repair? J Coll Physicians Surg Pak. 2011;21(5):291-296.
12. Schmedt CG, Sauerland S, Bittner R. Comparison of endoscopic procedures vs Lichtenstein and other open mesh techniques for inguinal hernia repair. Surg Endosc. 2005;19:188-199. doi: 10.1007/s00464-004-9126-0.
13. HerniaSurge Group. International guidelines for groin hernia management. Hernia. 2018;22(1):1-165. doi: 10.1007/s10029-017-1668-x.
14. Haladu N, Alabi A, Brazzelli M, et al. Open versus laparoscopic repair of inguinal hernia: an overview of systematic reviews of randomised controlled trials. Surg Endosc. 2022;36(7):4685-4700. doi: 10.1007/s00464-022-09161-6.
15. Patterson TJ, Beck J, Currie PJ, Spence RAJ, Spence G. Meta-analysis of patient-reported outcomes after laparoscopic versus open inguinal hernia repair. Br J Surg. 2019;106(7):824-836. doi: 10.1002/bjs.11139.
16. Pisanu A, Podda M, Saba A, Porceddu G, Uccheddu A. Meta-analysis and review of prospective randomized trials comparing laparoscopic and Lichtenstein techniques in recurrent inguinal hernia repair. Hernia. 2015;19(3):355-366. doi: 10.1007/s10029-014-1281-1.
17. Yang J, Tong da N, Yao J, Chen W. Laparoscopic or Lichtenstein repair for recurrent inguinal hernia: a meta-analysis of randomized controlled trials. ANZ J Surg. 2013;83(5):312-318. doi: 10.1111/ans.12010.
18. Ulusoy S, Özer M, Albuz Ö, Parlak Ö. Comparison of Lichtenstein and total extraperitoneal laparoscopic hernia repair (TEP) methods in inguinal hernias. Ortadogu Tıp Derg. 2019;11(4):366-371. doi: 10.21601/ortadogutipdergisi.464026.
19. Hernandez-Rosa J, Lo CC, Choi JJ, et al. Laparoscopic versus open inguinal hernia repair in octogenarians. Hernia. 2011;15(6):655-658. doi: 10.1007/s10029-011-0838-5.
20. Ciftci F. Laparoscopic versus open inguinal hernia repair on patients over 75 years of age Int J Clin Exp Med. 2015;8(6):10016-10020.
21. Lal P, Kajla RK, Chander J, Saha R, Ramteke VK. Randomized controlled study of laparoscopic total extraperitoneal versus open Lichtenstein inguinal hernia repair. Surg Endosc. 2003;17(6):850-856. doi: 10.1007/s00464-002-8575-6.

DOI: https://doi.org/10.18621/eurj.1661091 *Family Medicine*

Digital twins in personalized medicine: Insights from a systematic review

Ebru Uğraş

Department of Family Medicine, Ankara Yıldırım Beyazıt University School of Medicine, Ankara, Türkiye

ABSTRACT

Objectives: Digital twins (DTs) are increasingly recognized as valuable tools in the advancement of personalized medicine (PM).

Methods: This systematic review explores the current landscape of DT applications across medical domains, focusing on their feasibility, benefits, challenges, and potential future roles. Following PRISMA guidelines, literature was sourced from PubMed and Scopus using clearly defined inclusion and exclusion criteria, resulting in 14 studies eligible for full-text review.

Results: Findings show that DTs are being utilized in various specialties such as cardiology, geriatrics, radiology, and oncology, offering advantages in cost reduction, treatment precision, and patient outcomes. However, key challenges remain, including data privacy, technical integration, ethical considerations, and the need for interdisciplinary collaboration.

Conclusions: This systematic review offers a comprehensive synthesis of how DTs are being integrated into PM and highlights areas requiring further empirical research to support wider implementation.

Keywords: Digital twin, healthcare, personalized medicine, systematic review

Modern medicine is transitioning from a reactive treatment-based system to a proactive, interdisciplinary framework that targets personalized, systematic, and precise treatment plans [1]. Personalized medicine (PM), emerging conceptually in the late 20th century, adopts a blend of preventive, personalized, participatory, and predictive measures [2]. This approach molds healthcare services to fit the patient's biological markers, personal preferences, and environmental and genetic backgrounds, aiding physicians in choosing optimal treatments [3]. This customization is facilitated by diagnostics identifying patient-specific biological indicators, potentially revolutionizing care in areas such as cancer, Alzheimer's, cardiovascular diseases, and hepatitis [4]. The adoption of PM could significantly decrease healthcare costs and errors while enhancing the quality of care [5].

The digital twin (DT) concept was introduced in 2003 and involves using advanced physical models, sensor updates, and historical data to replicate the life span of its physical counterparts [6]. DTs consist of three elements: physical objects, their virtual replicas, and the interconnected data [7]. In healthcare, DTs offer a vision for integrating dynamic and consistent clinical data over time, utilizing statistical and mechanical models [8]. These models enhance clinical decision-making, increase interpretability, and lever-

Received: March 19, 2025 **Accepted:** July 21, 2025 **Available Online:** July 28, 2025 **Published:** September 4, 2025

How to cite this article: Uğraş E. Digital twins in personalized medicine: Insights from a systematic review. Eur Res J. 2025;11(5):1015-1022. doi: 10.18621/eurj.1661091.

Corresponding author: Ebru Uğraş, MD., Assist. Prof., Phone: +90 312 906 20 00, E-mail: ebruugras@hotmail.com

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

age big data for demographic analysis [9]. The application of DTs in healthcare is diverse, facilitating disease monitoring, diagnostic trials, and holistic health assessments [10, 11]. Despite the vast potential, DT implementation faces challenges including privacy concerns, data access, ethical issues, and potential biases [12]. This paper aims to provide a comprehensive overview of the utility, application, benefits, risks, and prospects of DTs in PM, highlighting how they can revolutionize personalized health monitoring, diagnostics, prognostics, prevention, and treatment [13].

Integrating DTs in PM is considered a natural and complementary strategy that can lead to pioneering technologies in the healthcare sector [14]. These technologies enable personalized healthcare solutions by leveraging real-time health data, thus enhancing patient outcomes through tailored treatment and intervention plans. Additionally, DTs facilitate the performance of virtual trials using computer models, which can simulate a variety of interventions from lifestyle changes to pharmaceutical regimens and surgical procedures [15]. This capability allows for the optimization of health strategies before they are applied in real-world settings, minimizing risks and maximizing effectiveness.

Moreover, comprehensive personalized DTs can incorporate a wide range of data including physical activities like sports and hobbies, mental activities such as thought processes and acquired knowledge, as well as vital organ metrics [16]. The use of DTs extends to specific medical fields such as diabetic modeling, pediatric cardiology, and artificial pancreas systems, demonstrating their broad applicability and utility [17, 18].

Despite this promise, the implementation of DTs in healthcare faces grouped sets of critical challenges. These challenges can be categorized as follows:

1. Technical Limitations: These include insufficient model fidelity, challenges in integrating real-time data, and limited validation across different clinical settings.

2. Medical Uncertainties: Concerns regarding diagnostic accuracy, generalizability of simulations, and lack of standardized clinical protocols.

3. Ethical and Privacy Concerns: Risks related to data privacy, consent procedures, algorithmic bias, and potential for discrimination [19].

The future of DTs in personalized medicine is promising, with ongoing research and advancements enhancing their accuracy and functionality. Emerging technologies are improving real-time data processing, enabling dynamic and precise health monitoring. This is vital for conditions requiring rapid responses, like cardiac emergencies or acute infections. Additionally, DTs are expanding beyond individual care to assist in population health management by simulating disease spread, helping public health officials craft more effective strategies [20].

To address these grouped challenges, several strategic solutions are proposed:

1. Technological Advancements: Innovations in data processing, AI integration, and secure communication frameworks aim to improve the accuracy and reliability of DTs.

2. Interdisciplinary Collaboration: Cooperation among clinicians, data scientists, and ethicists is seen as essential to guide ethical implementation and maximize clinical benefit.

3. Education and Awareness: Educating healthcare professionals and the public can help mitigate skepticism, build trust, and promote informed adoption of DT-based technologies [17, 19].

The primary aim of this research is to thoroughly explore the capabilities and scope of DTs within the context of PM. This study intends to assess both the current applications and the potential future roles of DTs, analyzing how they can enhance the personalization and efficacy of healthcare treatments. By providing a detailed review of the integration of DTs in PM, the research seeks to understand how these digital replicas can optimize health outcomes through individualized treatment plans and proactive health management [20, 21].

The importance of this study lies in its potential to revolutionize healthcare delivery. DTs offer a promising avenue for advancing personalized medicine by enabling a more nuanced understanding of individual health conditions through the simulation of various medical interventions and outcomes. This approach not only helps in refining treatment protocols but also in anticipating potential complications before they arise, thus contributing to safer, more effective healthcare practices. Furthermore, as the healthcare industry continues to evolve towards more data-driven and patient-centered models, the insights garnered from this research could significantly influence future health-

care policies and practices, making the adoption of DTs a critical area of investigation [22, 23].

METHODS

This study aimed to investigate the role of DTs in PM through a systematic review. The goal was to evaluate their feasibility, practical applications, benefits, and limitations across various medical disciplines. The review covered literature published from 1999 through April 1, 2024.

The review process adhered to the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines. Two electronic databases - PubMed and Scopus - were searched comprehensively. The search approach involved a structured combination of terms relevant to digital twins and personalized medicine. No date limits were initially applied to ensure comprehensive coverage.

Eligibility Criteria

Studies were included if they met the following criteria: (1) Focused on digital twin technologies applied within the scope of personalized medicine; (2) Described human-centered digital twin models; (3) Included a clearly stated methodology; and (4) Targeted a specific patient group or disease category.

Excluded works were: (1) Not written in English, (2) Not freely accessible in full-text format, and (3) Classified as books, book chapters, letters, editorials, or conference proceedings.

Screening and Selection Process

An initial pool of 400 articles was identified through the database searches. After removing duplicates, 315 unique articles remained. Titles and abstracts were screened for relevance to the research topic. After this initial screening, 60 articles were retained for full-text evaluation. Based on a detailed application of the inclusion and exclusion criteria, 14 studies were selected for inclusion in the final synthesis. The PRISMA flowchart (Fig. 1) summarizes this process.

Data Extraction and Quality Control

Data extraction was independently carried out by two reviewers. Key details captured from each study included country of origin, year of publication, medical domain, population characteristics, type of DT model, and reported outcomes. Any disagreements during the extraction process were resolved through discussion or consultation with a third reviewer to ensure consistency.

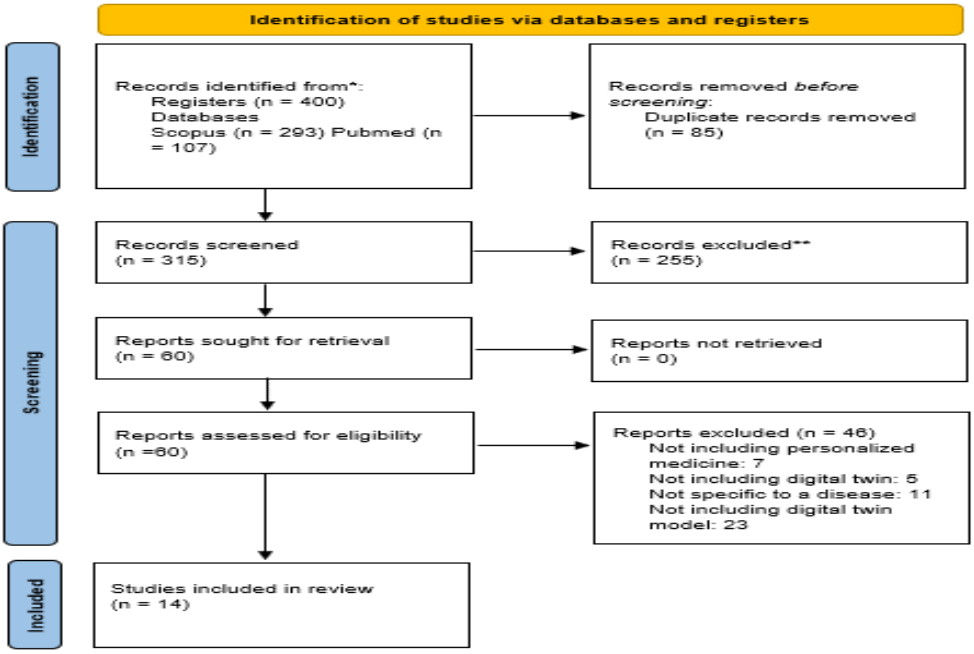

Fig. 1. PRISMA flow diagram.

Assessment of Bias and Limitations

No formal risk of bias tool was used, due to the descriptive and exploratory nature of the included studies. However, any reported methodological limitations within the selected studies were noted and are discussed in the relevant sections.

This approach provides a transparent and reproducible method for evaluating the current evidence surrounding digital twins in personalized medicine, with careful attention to study selection, data handling, and reporting integrity.

Statistical Analysis

The common themes of the studies were classified qualitatively, and the findings were analyzed descriptively. The areas of application of digital twin technologies in personalized medicine, their clinical contributions, and the challenges encountered were evaluated based on frequency using the number of studies. No statistical software was used in this analysis process; basic data editing and analysis were performed using standard methods.

RESULTS

This section synthesizes the findings from 14 studies that met the inclusion criteria. The review presents a descriptive analysis of study characteristics, clinical domains, thematic insights, and implementation trends, allowing for a consolidated view of how digital twins (DTs) are being applied across PM (Table 1).

Study Characteristics and Distribution

Studies were conducted between 2019 and 2023, with an upward trend in recent years—especially in 2022 and 2023. Geographically, Germany contributed the most individual studies (3), followed by collaborative research involving institutions from the USA, Canada, UK, and Australia. The international nature of the studies indicates broad academic interest and growing investment in DTs across various health systems.

Clinical Application Areas

The studies addressed diverse medical fields:

1. Cardiology (3 studies): Simulations of atrial fibrillation, ablation strategies, and cardiac electrophysiology modeling.

2. Geriatrics (3 studies): Management of dementia, diabetes, and elderly care modeling.

3. Radiology and Oncology (3 studies): Imaging enhancement and theranostic modeling.

4. Others (5 studies): Neurology, gastroenterology, nutrition, diabetes, and surgery.

This distribution highlights the potential scalability of DTs across a broad spectrum of clinical needs.

Aggregated Findings and Evidence Patterns

A synthesis of core themes reported across the literature yielded the following:

1. Adoption Benefits: Improvement in individualized treatment planning and risk prediction (reported in 12/14 studies). Reduced trial-and-error in treatment decisions through virtual testing (reported in 10/14 studies). Support for non-invasive, real-time patient monitoring (reported in 9/14 studies).

2. Challenges Identified: Patient data privacy and protection were major concerns in 9 studies. Interoperability limitations and lack of unified data standards reported in 7 studies. Algorithmic bias and ethical transparency concerns were addressed in 6 studies.

3. Technical Innovations: Machine learning and AI integration to enhance DT model performance (5/14). Use of cloud-based infrastructure and wearable devices for real-time feedback (4/14). Multi-scale modeling to integrate physiological, behavioral, and genomic data (3/14).

4. Future Directions and Recommendations: Need for standardized validation frameworks and regulatory oversight (recommended in 6 studies). Calls for more clinical trials or pilot programs to demonstrate real-world feasibility (recommended in 5/14 studies). Emphasis on user-centered design to improve usability for both clinicians and patients (recommended in 4/14 studies).

Presentation and Synthesis Style

Consistent with systematic review methodology, all interpretations are based on descriptive aggregation and thematic coding of reported results. Quantitative counts reflect frequencies of themes across the included studies, offering clarity without over-interpreting underlying evidence.

Table 1. Overview of analyzed studies

Author	Country-Year	Population	Field of personalized medicine
Currie *et al.* [16]	USA, Australia - 2023	-	Radiology
Thamotharan *et al.* [24]	USA, Italy - 2023	15 patients	Geriatrics
Abeltino *et al.* [25]	Italy - 2023	10 users tested	Nutrition
Azzolin *et al.* [26]	Germany - 2023	29 patient model	Cardiology
Rahmim *et al.* [27]	USA, Canada - 2022	-	Oncological radiology
Wickramasinghe *et al.* [28]	Australia - 2022	-	Geriatrics/Dementia
Sahal *et al.* [29]	Ireland, Yemen - 2022	-	Health sector
Pinton [30]	Denmark - 2022	-	General surgery and Gastroenterology
Jung *et al.* [31]	Austria - 2022	7 individuals	Cardiology
Barbiero *et al.* [32]	UK - 2021	2 clinical cases	Health sector/Whole human body
Voight *et al.* [33]	Germany - 2021	-	Neurology
Geissler *et al.* [6]	Germany - 2021	200 records used	Radiology
Corral-Acero *et al.* [20]	UK, Belgium - 2020	-	Cardiology
Liu *et al.* [15]	USA, Canada, China - 2019	Case studies	Geriatrics

This section provides a cohesive and integrated overview of the status of digital twin technologies in personalized medicine, serving as a foundation for discussion and further exploration in subsequent sections.

DISCUSSION

The studies reviewed indicate that Germany is the single country with the highest number of individual studies, while the United States featured prominently in multinational studies. The majority of research occurs in developed countries such as the USA, Germany, Italy, and the United Kingdom, suggesting that these nations are focusing significantly on advancing technologies in the field of PM using DTs. The surge in publications in 2022 and 2023 compared to other years suggests that the use of DTs in PM is a relatively new and increasingly popular area of study [34].

Most research in the field has been conducted in cardiology and geriatrics. Considering that 33.42% of mortality in Turkey is due to circulatory system diseases in 2023, the application of DTs in cardiology is particularly critical [35]. DTs developed for circulatory diseases could potentially reduce or even eliminate mortality caused by these conditions [36]. Additionally, with the life expectancy in OECD countries at 80.3 years, 79 years in Turkey, and the global average at 71 years, the aging population is significant [37]. In Türkiye, the elderly population grew by 22.5% in five years, reaching over 8 million in 2017. By 2040, this figure is expected to rise to 16.3% and 25.6% by 2080 [35]. DTs could play a transformative role in reducing age-related ailments, effectively 'rejuvenating' the concept of aging.

One of the studies used a DT approach in a precision treatment program for 64 patients with type-2 diabetes over 90 days, observing significant reductions in diastolic and systolic blood pressure, glycemic variability, and body mass index, with nearly all patients ceasing to use antihypertensive drugs [10].Furthermore, Digital Health Twins (DHTs) can facilitate bed planning, staff schedules, early disease diagnosis, treatment trials, surgical simulations, and virtual drug trials, potentially detecting errors early, identifying diseases before they occur, reducing risks, and saving costs [12].

Huang *et al*. developed an ethical framework to

examine the ethical issues in using DTs for PM across four stages: data collection, data management, data analysis, and information use. Each stage presents unique challenges, including privacy concerns, autonomy, informed consent deficiencies, biases in algorithmic decision-making, and potential misuse of data that could distort healthcare understanding and lead to discrimination or health inequities [38].

The use of internet-connected wearable devices exposes individuals to the risks of cybersecurity threats, identity theft, fraud, and data breaches. It also raises concerns about discrimination against people with pre-existing health conditions, potentially affecting insurance coverage or job opportunities. Moreover, self-monitoring through these devices could lead to obsessive behaviors regarding health [39].

Overall, DTs for PM are proving their utility and effectiveness across various medical fields, including geriatrics, dementia, radiology, cardiology, oncology, diabetes, neurology, nutrition, and the broader health sector. Future research should focus on expanding the technology and testing environments to overcome the complexities and ethical challenges posed by this innovative approach. The potential benefits for patient care and health system economics are immense, suggesting a significant shift towards more personalized, efficient, and preventative healthcare solutions.

Limitations

The fact that digital twin technology is still an emerging field limits the generalizability of findings due to the lack of standardized definitions and measurement methods. Therefore, more comprehensive, controlled, and long-term studies in the future will increase the knowledge base in this field. The fact that digital twin technology is still an emerging field limits the generalizability of findings due to the lack of standardized definitions and measurement methods. Therefore, more comprehensive, controlled, and long-term studies in the future will increase the knowledge base in this field.

CONCLUSION

The systematic review of Digital Twins (DTs) in personalized medicine (PM) highlights their growing impact across various medical fields, showcasing their potential to transform healthcare. Our findings reveal significant DT adoption in developed countries, with a concentration of research in Germany and the United States, indicating a close connection between DT development and advanced medical research infrastructures focused on innovative healthcare technologies. The rise in research publications in 2022 and 2023 highlights the growing importance of Digital Twins (DTs) in personalized medicine (PM), particularly in improving diagnostic accuracy and treatment personalization. Studies show that DTs effectively address complex health issues, especially in cardiology and geriatrics, which are critical as the global population ages. Additionally, DTs have shown promise in managing chronic diseases like diabetes by reducing key health indicators such as blood pressure and glycemic variability, potentially lowering the need for routine medications and offering cost savings in long-term management.

The integration of Digital Twins (DTs) into clinical practice faces challenges, especially regarding ethics, patient autonomy, privacy, and data security. The handling of large health data raises concerns about privacy and misuse, requiring strict regulations to protect patient information. Additionally, unequal access to DTs may worsen health disparities unless technology is more widely accessible. Addressing these issues demands a collaborative approach from technology, medicine, ethics, and policy. In conclusion, DTs offer great potential for personalized medicine, but future research must address technical and ethical obstacles, improve system interoperability, and broaden their use across medical fields to enhance healthcare outcomes and efficiency.

Ethical Statement

As this study is a systematic review, ethical committee approval is not required. There are no human or animal elements in our study. Data were obtained from open sources on the internet.

Data Availability

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

Authors' Contribution

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

Conflict of interest

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

Financing

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

Generative Artificial Intelligence Statement

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

Editor's note

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

REFERENCES

1. Yuan B. What Personalized Medicine Humans Need and Way to It --also on the Practical Significance and Scientific Limitations of Precision Medicine. Pharmgenomics Pers Med. 2022;15:927-942. doi: 10.2147/PGPM.S380767.
2. Mishra V, Chanda P, Tambuwala MM, Suttee A. Personalized medicine: An overview. Int J Pharm Qual Assur. 2019;10(2):290-294. doi: 10.25258/ijpqa.10.2.13.
3. Pritchard DE, Moeckel F, Villa MS, Housman LT, McCarty CA, McLeod HL. Strategies for integrating personalized medicine into healthcare practice. Per Med. 2017 ;14(2):141-152. doi: 10.2217/pme-2016-0064.
4. Ginsburg GS, Willard HF. Genomic and personalized medicine: foundations and applications. Transl Res. 2009;154(6):277-287. doi: 10.1016/j.trsl.2009.09.005.
5. Calcaterra V, Pagani V, Zuccotti G. Digital Twin: A Future Health Challenge in Prevention, Early Diagnosis and Personalisation of Medical Care in Paediatrics. Int J Environ Res Public Health. 2023;20(3):2181. doi: 10.3390/ijerph20032181.
6. Geissler F, Heiß R, Kopp M, et al. Personalized computed tomography - Automated estimation of height and weight of a simulated digital twin using a 3D camera and artificial intelligence. Rofo. 2021;193(4):437-445. doi: 10.1055/a-1253-8558.
7. Shamanna P, Dharmalingam M, Sahay R, et al. Retrospective study of glycemic variability, BMI, and blood pressure in diabetes patients in the Digital Twin Precision Treatment Program. Sci Rep. 2021;11(1):14892. doi: 10.1038/s41598-021-94339-6.
8. de Sio FS, Faber NS, Savulescu J, Vincent NA. Why less praise for enhanced performance? Moving beyond responsibility-shifting, authenticity, and cheating toward a natu re-of-activities approach. In: Jotterand F, Dubljevic V, editors. Cognitive Enhancement: Ethical and Policy Implications in International Perspectives. New York: Oxford Academic. 2016. pp. 27-41.
9. Francis F. Our Posthuman Future: Consequences of the Biotechnology Revolution. Profile Books; 2003.
10. Chu Y, Li S, Tang J, Wu H. The potential of the Medical Digital Twin in diabetes management: a review. Front Med (Lausanne). 2023;10:1178912. doi: 10.3389/fmed.2023.1178912.
11. Hernandez-Boussard T, Macklin P, Greenspan EJ, et al. Digital twins for predictive oncology will be a paradigm shift for precision cancer care. Nat Med. 2021;27(12):2065-2066. doi: 10.1038/s41591-021-01558-5.
12. Spitzer M, Dattner I, Zilcha-Mano S. Digital twins and the future of precision mental health. Front Psychiatry. 2023;14:1082598. doi: 10.3389/fpsyt.2023.1082598.
13. Misra SC, Bisui S, Singh A. A study on the role of trust factor in adopting personalized medicine. Behav Inf Technol. 2020;39(7):771-787. doi: 10.1080/0144929X.2019.1614225.
14. Glaessgen EH, Stargel DS. The Digital Twin Paradigm for Future NASA and U.S. Air Force Vehicles. In: Paper for the 53rd Structures, Structural Dynamics, and Materials Conference: Special Session on the Digital Twin. American Institute of Aeronautics and Astronautics; 2012.
15. Liu Y, Zhang L, Yang Y, et al. A Novel Cloud-Based Framework for the Elderly Healthcare Services Using Digital Twin. IEEE Access. 2019;7:49088-490101. doi: 10.1109/ACCESS.2019.2909828.
16. Currie GM, Rohren EM. Radiation Dosimetry, Artificial Intelligence and Digital Twins:Old Dog, New Tricks. Semin Nucl Med. 2023;53(3):457-466. doi: 10.1053/j.semnuclmed.2022.10.007.
17. Bruynseels K, Santoni de Sio F, van den Hoven J. Digital Twins in Health Care: Ethical Implications of an Emerging Engineering Paradigm. Front Genet. 2018;9:31. doi: 10.3389/fgene.2018.00031.
18. Alber M, Buganza Tepole A, Cannon WR, et al. Integrating machine learning and multiscale modeling-perspectives, challenges, and opportunities in the biological, biomedical, and behavioral sciences. NPJ Digit Med. 2019;2:115. doi: 10.1038/s41746-019-0193-y.
19. Iqbal JD, Krauthammer M, Biller-Andorno N. The Use and Ethics of Digital Twins in Medicine. J Law Med Ethics. 2022;50(3):583-596. doi: 10.1017/jme.2022.97.
20. Corral-Acero J, Margara F, Marciniak M, et al. The 'Digital Twin' to enable the vision of precision cardiology. Eur Heart J. 2020;41(48):4556-4564. doi: 10.1093/eurheartj/ehaa159.
21. Gillette K, Gsell MAF, Bouyssier J, et al. Automated Frame-

work for the Inclusion of a His-Purkinje System in Cardiac Digital Twins of Ventricular Electrophysiology. Ann Biomed Eng. 2021;49(12):3143-3153. doi: 10.1007/s10439-021-02825-9.
22. Schmidt A, Helgers H, Vetter FL, Juckers A, Strube J. Digital twin of mRNA-based SARS-COVID-19 vaccine manufacturing towards autonomous operation for improvements in speed, scale, robustness, flexibility and real-time release testing. Processes. 2021;9(5):748. doi: 10.3390/pr9050748.
23. Laubenbacher R, Sluka JP, Glazier JA. Using digital twins in viral infection. Science. 2021;371(6534):1105-1106. doi: 10.1126/science.abf3370.
24. Thamotharan P, Srinivasan S, Kesavadev J, et al. Human Digital Twin for Personalized Elderly Type 2 Diabetes Management. J Clin Med. 2023;12(6):2094. doi: 10.3390/jcm12062094.
25. Abeltino A, Bianchetti G, Serantoni C, Riente A, De Spirito M, Maulucci G. Putting the Personalized Metabolic Avatar into Production: A Comparison between Deep-Learning and Statistical Models for Weight Prediction. Nutrients. 2023;15(5):1199. doi: 10.3390/nu15051199.
26. Azzolin L, Eichenlaub M, Nagel C, et al. Personalized ablation vs. conventional ablation strategies to terminate atrial fibrillation and prevent recurrence. Europace. 2023;25(1):211-222. doi: 10.1093/europace/euac116.
27. Rahmim A, Brosch-Lenz J, Fele-Paranj A, et al. Theranostic digital twins for personalized radiopharmaceutical therapies: Reimagining theranostics via computational nuclear oncology. Front Oncol. 2022;12:1062592. doi: 10.3389/fonc.2022.1062592.
28. Wickramasinghe N, Ulapane N, Andargoli A, et al. Digital twins to enable better precision and personalized dementia care. JAMIA Open. 2022;5(3):ooac072. doi: 10.1093/jamiaopen/ooac072.
29. Sahal R, Alsamhi SH, Brown KN. Personal Digital Twin: A Close Look into the Present and a Step towards the Future of Personalised Healthcare Industry. Sensors (Basel). 2022;22(15):5918. doi: 10.3390/s22155918.
30. Pinton P. Computational models in inflammatory bowel disease. Clin Transl Sci. 2022;15(4):824-830. doi: 10.1111/cts.13228. Epub 2022 Feb 5.
31. Jung A, Gsell MAF, Augustin CM, Plank G. An Integrated Workflow for Building Digital Twins of Cardiac Electromechanics-A Multi-Fidelity Approach for Personalising Active Mechanics. Mathematics (Basel). 2022;10(5):823. doi: 10.3390/math10050823.
32. Barbiero P, Viñas Torné R, Lió P. Graph Representation Forecasting of Patient's Medical Conditions: Toward a Digital Twin. Front Genet. 2021;12:652907. doi: 10.3389/fgene.2021.652907.
33. Voigt I, Inojosa H, Dillenseger A, Haase R, Akgün K, Ziemssen T. Digital Twins for Multiple Sclerosis. Front Immunol. 2021;12:669811. doi: 10.3389/fimmu.2021.669811.
34. Eumi EM. A systematic review of Digital Twins in efficient pandemic management with challenges and emerging trends. Decision Analytics J. 2024;12:100502. doi: 10.1016/j.dajour.2024.100502.
35. TÜİK. Ölüm ve Ölüm Nedeni İstatistikleri, 2023 [Internet] https://data.tuik.gov.tr/Bulten/Index?p=Olum-ve-Olum-Nedeni-Istatistikleri-2023-53709 (Accessed: 01.12.224)
36. Sel K, Osman D, Zare F, et al. Building Digital Twins for Cardiovascular Health: From Principles to Clinical Impact. J Am Heart Assoc. 2024;13(19):e031981. doi: 10.1161/JAHA.123.031981.
37. OECD, Society at a Glance 2024: OECD Social Indicators, OECD Publishing, Paris. 2024. doi: 10.1787/918d8db3-en.
38. Huang PH, Kim KH, Schermer M. Ethical Issues of Digital Twins for Personalized Health Care Service: Preliminary Mapping Study. J Med Internet Res. 2022;24(1):e33081. doi: 10.2196/33081.
39. Thapa S, Bello A, Maurushat A, Farid F. Security Risks and User Perception towards Adopting Wearable Internet of Medical Things. Int J Environ Res Public Health. 2023;20(8):5519. doi: 10.3390/ijerph20085519.

The European Research Journal 2025;11(5):1023-1028 Case Report

DOI: https://doi.org/10.18621/eurj.1668293 *Pediatric Surgery*

A rare disease presenting with acute abdominal pain in a girl: Solitary cecal diverticulum

Hakan Özcan[1], Mehmet Özer[2], Mete Kaya[1]

[1]Department of Pediatric Surgery, University of Health Sciences, Bursa Yüksek Ihtisas Training and Research Hospital, Bursa, Türkiye;
[2]Department of Medical Patalogy, University of Health Sciences, Bursa Yüksek Ihtisas Training and Research Hospital, Bursa, Türkiye

ABSTRACT

Solitary cecal diverticulum (SCD) is a rare gastrointestinal disease, especially in children, predominantly observed in middle-aged Asian men, and presented with nonspecific symptoms similar to acute appendicitis. Herein we report a case of SCD in a 15-year-old girl who had acute abdominal pain symptoms, and was diagnosed intraoperatively. The SCD, which contained a fecalith, was excised, followed by cecal repair and an appendectomy. Histopathological examination confirmed the diagnosis of SCD. This case highlights the importance of considering SCD in the differential diagnosis of acute abdominal pain, particularly in pediatric patients.

Keywords: Solitary cecal diverticulum, acute abdominal pain, diverticulitis, fecalith

Solitary cecal diverticulum (SCD), first described by Potier in 1912, is a rare gastrointestinal disease (1). SCD has been reported to constitute only 3.6% of all colonic diverticula (2). It has been reported that the disease typically manifests in the fifth decade of life and is more commonly observed in Asian men. SCD often presents with nonspecific symptoms such as right lower quadrant abdominal pain, vomiting, and loss of appetite, making its differentiation from acute appendicitis challenging (3). It is rarely seen in the pediatric population, with only a limited number of case series or reports available in the literature (4-7). In this study, we present a case of SCD in an adolescent girl who presented with acute abdominal pain due to fecalith-induced inflammation.

CASE PRESENTATION

A 15-year-old female patient with no previous medical history was admitted to the emergency department with the sudden onset of abdominal pain and vomiting persisting for two days. Vital signs were within normal limits. Physical examination revealed tenderness, guarding, and rebound tenderness in the right lower quadrant of the abdomen. Laboratory results showed a white blood cell count of 11,510×10^3/mL, a left shift (Neutrophils: 74.3%), and elevated C-reactive protein (17.4 mg/L). An upright abdominal X-ray did not reveal any diagnostic findings. Abdominal ultrasonography (USG) demonstrated edema in the ileocecal loop and adjacent mesentery in the right lower quad-

Received: March 30, 2025 **Accepted:** June 14, 2025 **Available Online:** June 17, 2025 **Published:** September 4, 2025

How to cite this article: Özcan H, Özer M, Kaya M. A rare disease presenting with acute abdominal pain in a girl: Solitary cecal diverticulum. Eur Res J. 2025;11(5):1023-1028. doi: 10.18621/eurj.1668293

Corresponding author: Hakan Özcan, MD., Phone: +49 40 7410 - 57228, E-mail: drhknzcn@gmail.com

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

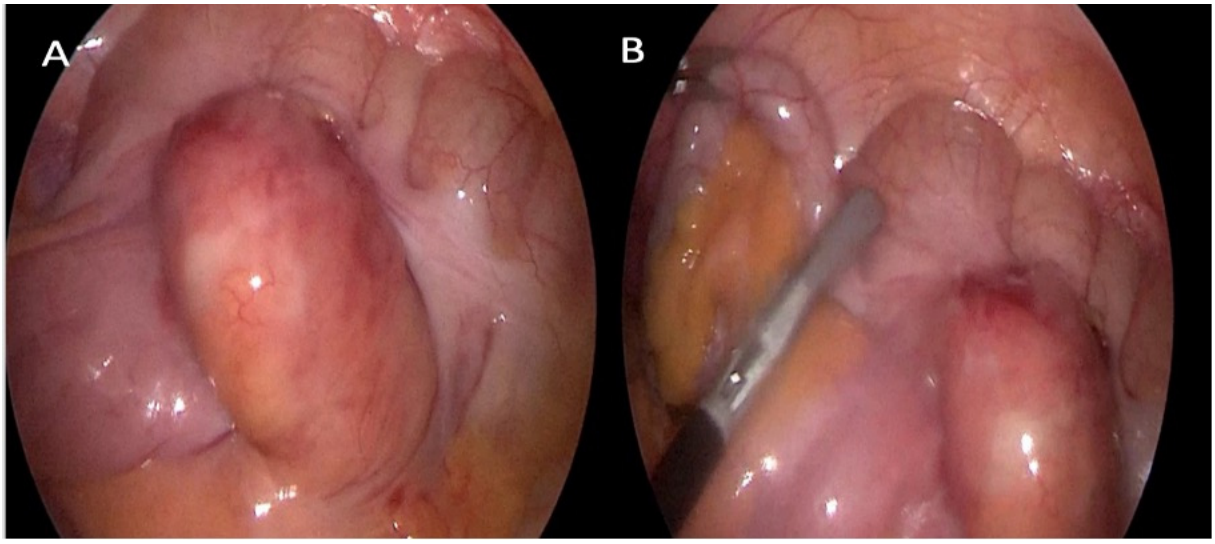

Fig. 1. A) Inflamed mass (SCD) located on the anterior wall of the cecum adjacent to the ileocecal junction during laparoscopic examination. B) Normal appearance of the appendix.

rant, along with approximately 1 cm of free fluid in the Douglas pouch. In addition, a blind-ending luminal structure containing a fecalith was observed on the anterior surface of the cecum. After initial resuscitation, laparoscopy was performed. On laparoscopy, an inflamed, pedunculated mass was identified on the anterior wall of the cecum, while the appendix appeared normal (Fig. 1A and Fig. 1B). When the mass was excised from its neck, it was noted to be a luminal structure containing a fecalith. To ensure safer repair, conversion to laparotomy was performed. Diverticulectomy and bowel wall repair in two layers were

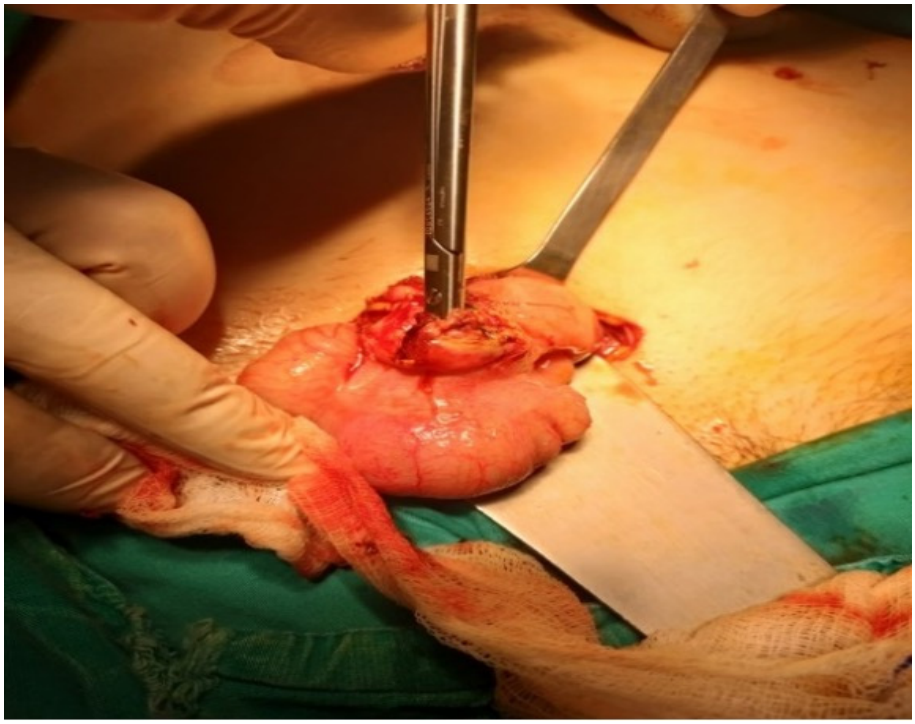

Fig. 2. The connection between the SCD and the cecal lumen is observed during laparotomy.

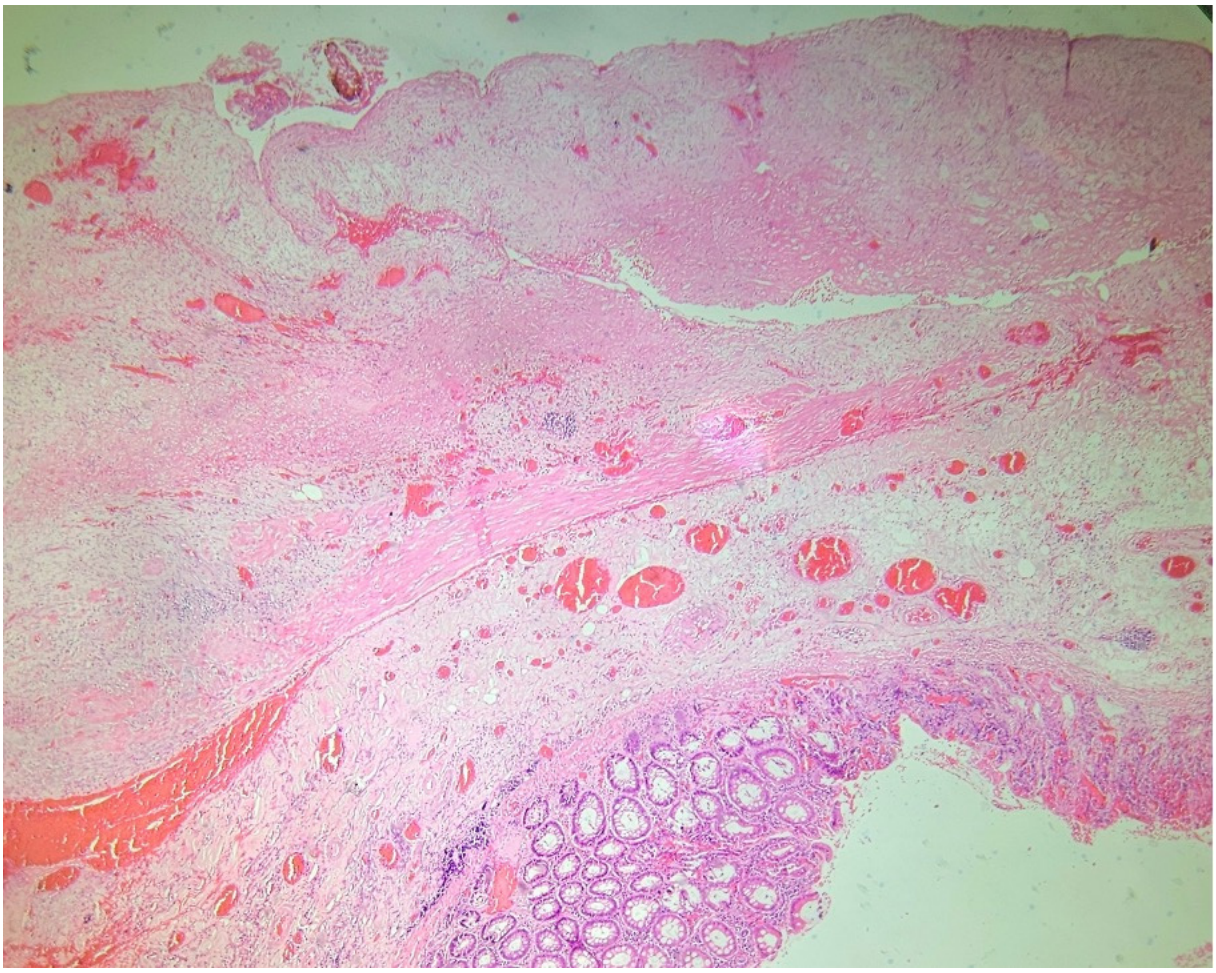

Fig. 3. Histopathologic examination revealed that the SCD structure was lined with colonic epithelium and involved all layers of the bowel, exhibiting vascular dilation, congestion, and acute inflammation within its wall (Hematoxylin-eosin, ×40).

performed (Fig. 2). The procedure was concluded after formal appendectomy. The postoperative course was uneventful. Histopathologic examination revealed colonic tissue involving all bowel layers, exhibiting vascular dilation, congestion, and acute inflammation, and suggested an SCD association with diverticulitis (Fig. 3). The patient was followed up for two years without complications.

DISCUSSION

Colonic diverticula are commonly observed in older age groups and are rarely symptomatic. The etiology of the condition has been linked to a sedentary lifestyle, obesity, and a diet deficient in fiber. It has been suggested that diverticula are more frequently observed in the left colon due to increased intraluminal pressure, and the disease is considered an acquired pathology (8). On the other hand, the incidence of right colonic diverticula has been reported to increase in conditions such as connective tissue disorders (e.g., Marfan syndrome, Ehlers-Danlos syndrome, Williams-Beuren syndrome) and neuroenteric anomalies like intestinal hypoganglionosis (9). These types of diverticula may be located throughout the colon and may be numerous. Because they are contained within the mucosa-submucosa without involvement of the muscularis layer, it is still debated whether they are true diverticula. SCD is considered distinct from other colonic diverticula due to its location and histopathologic properties, which involve all bowel layers (10). One theory proposes that SCD arises from a persistent "temporary appendix" structure, which develops dur-

ing the first six weeks of the embryologic period and fails to regress as expected. SCDs are typically located approximately 2-3 cm from the ileocecal valve and are positioned anteriorly (11). In our case, the diverticulum was found on the anterior surface of the cecum, attached via a narrow neck similar to the appendix. The histopathological examination revealed that the lesion involved all layers of the normal colonic wall, and we believe this supports the congenital malformation theory of SCD etiology mentioned above.

Since its first description, up to 1,000 cases of SCD have been reported in the literature across various age groups (12). In a study by Sardi *et al*. (2), which included 881 adult cases, SCD constituted 3.6% of all colonic diverticula, with an average age of 43.6 years. It was noted that colonic diverticula are more frequently encountered in men compared to women. In a systematic review, data were collected from a total of 988 patients, aged 20 to 73 years. The male-to-female ratio in this study was reported to be approximately 1.12:1 (13). In the pediatric population, it is difficult to estimate the incidence and the male-to-female ratio due to its rarity. To date, 18 cases have been reported, with most of them being case reports (4-7).

SCDs remain asymptomatic unless complicated, and most of them can be detected incidentally during surgery. Symptomatic cases most commonly present with acute abdominal symptoms due to diverticulitis, bleeding, or perforation (14). Diagnosis of SCD, especially distinguishing it from appendicitis, can be difficult (2). In a systematic review of adult series, the most common symptoms of SCD were right lower quadrant abdominal pain, nausea and/or vomiting, and fever, with frequencies of 93.2%, 35.4%, and 26.9%, respectively (13). In our case, the main complaints were acute abdominal pain, nausea, vomiting, and loss of appetite, which mimicked acute appendicitis. Even in symptomatic cases, preoperative diagnosis of SCD through imaging is challenging (15).

In the literature, 22.8% of SCD patients were diagnosed via radiological imaging, while the remainder were diagnosed during surgery (13). In a study of 19 patients diagnosed with SCD, Wyble *et al*. (14) reported that only 2 (11%) were diagnosed preoperatively with detailed imaging, but these cases had previously undergone appendectomy, which allowed for the diagnosis to be made. On the other hand, in a prospective study by Chou *et al*. (16), involving 934 patients presenting with right lower quadrant pain, it was emphasized that USG is a more effective technique for differentiating SCD from appendicitis, with a sensitivity of 91.3% and specificity of 99.5%. In studies involving adults, the detection of a segmentally thickened colonic wall with an oval, hypoechoic structure protruding from the right colon on US has been suggested as an indicator of uncomplicated right colonic diverticula (17). Computerized tomography (CT) has also been used to diagnose SCD. If complicated, some characteristic appearances, such as thickening of the right colon wall, inflammation in the surrounding fat tissue, abscess formation, and local free air, may be diagnostic. However, these nonspecific imaging findings may also be seen in other pathologies, particularly in cecal carcinoma (18).

In pediatric SCD cases and series, no typical findings have been reported, and it is often incidentally discovered during surgery for acute abdominal conditions, especially appendicitis (4-7). In our case, preoperative USG suggested appendicitis with a fecalith. As is known, when a fecalith is found on imaging studies, acute appendicitis is typically diagnosed definitively. We also diagnosed our patient with acute appendicitis. This is the second case report of a pediatric patient presenting with diverticulitis caused by fecalith-induced inflammation.

Colonic diverticula, commonly encountered in the adult patient population, have established treatment standards. In uncomplicated cases, a conservative approach involving bowel rest and intravenous antibiotics is preferred, while in complicated and symptomatic cases, more aggressive surgeries such as diverticulectomy, ileocolic resection, and right hemicolectomy are performed (19). On the other hand, the recurrence rates of uncomplicated SCD during follow-up have been reported to range from 0% to 25%. For cases with frequent recurrences or complicated SCD, surgical resection is undoubtedly the most appropriate treatment option (20). The choice of surgical procedure depends on the location of the diverticulum and whether it is complicated. In cases with limited inflammation around the diverticulum, diverticulectomy with primary repair and appendectomy is recommended, while ileocolic resection or right hemicolectomy is recommended for complicated cases (21). Lane *et al*. (19) have recommended right hemicolectomy as the first-line treatment for adult SCD cases.

In their study, the mortality rate for those who underwent right hemicolectomy was found to be 18%, and they concluded that hemicolectomy should be reserved for complicated cases. Diverticulectomy can be performed by laparotomy or laparoscopy depending on experience. Due to technical difficulties, prolonged operative time, and increased anesthesia risks, laparoscopic surgery may not be suitable for complicated cases (22, 23).

Because of pediatric SCD series are limited, there is no established standard of the management of SCD. In children with SCD, it is reported that, instead of aggressive approaches, the less aggressive surgical methods for treatment of SCD are preferred (4-7). In our presented case, due to the limited inflammation around the diverticulum, diverticulectomy was preferred, and laparoscopy was converted to open surgery for safe excision. In seven (39%) of the reported pediatric SCD cases from four studies, conservative treatment was performed as the first-line therapy. Among these patients, conservative management was successful in four cases, while two patients subsequently required surgical intervention. Additionally, in one patient with a fecalith detected within the SCD, diverticulitis was treated by colonoscopic removal of the fecalith (4-7).

Although conservative management of uncomplicated SCD is considered the first-line treatment in adult series, surgery may be mandatory when treatment failure and recurrence occur. In various studies, no recurrences or major complications have been reported during follow-up in patients who underwent either simple diverticulectomy or hemicolectomy (19-23). In our case, no recurrences or complications were observed during the two-year follow-up period.

CONCLUSION

SCD is a rare condition in children. Due to the absence of typical signs and symptoms or imaging characteristics, it is often detected during acute abdominal surgery. There is no standardized approach to its treatment. When symptomatic or incidentally detected, simple excision is a safe and effective treatment for SCD.

Ethical Statement

Ethics Committee approval is not required for this study. This study is a case report.

Data Availability

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

Patient' Consent

Patient was informed about the purpose of the case report, and written informed consent was obtained from the patient's family for this publication.

Authors' Contribution

Study Conception: HÖ; Study Design: HÖ; Supervision: MK; Funding: M/A; Materials: HÖ, MÖ; Data Collection and/or Processing: HÖ, MÖ; Statistical Analysis and/or Data Interpretation: HÖ, MÖ; Literature Review: HÖ; Manuscript Preparation: HÖ, MÖ, MK; and Critical Review: MK.

Conflict of interest

The authors disclosed no conflict of interest during the preparation or publication of this manuscript.

Financing

The authors disclosed that they did not receive any grant during the conduction or writing of this study.

Generative Artificial Intelligence Statement

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

Editor's note

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

REFERENCES

1. Potier F. Diverticulite et appendicite. Bull Mem Soc Anat Paris.

1912;37:29-31.
2. Sardi A, Gokli A, Singer JA. Diverticular disease of the cecum and ascending colon. A review of 881 cases. Am Surg. 1987;53(1):41-45.
3. Cristaudo A, Pillay P, Naidu S. Caecal diverticulitis: Presentation and management. Ann Med Surg (Lond). 2015;4(1):72-75.
4. Lee ZW, Albright EA, Brown BP, Markel TA. Congenital cecal diverticulitis in a pediatric patient. J Pediatr Surg Case Rep. 2021;72:101929. doi: 10.1016/j.epsc.2021.101929.
5. Li X, Liu W, Zhang S, Liu H. Diagnosis and treatment experience of cecal diverticulitis in six pediatric patients. Front Pediatr. 2024;12:1478296. doi: 10.3389/fped.2024.1478296.
6. Abuelazm S, Kanumalla H, Taborda-Alvarez M, Eisdorfer J, Basak R. Acute Right-Sided Abdominal Pain in an Adolescent: Expect the Unexpected. Cureus. 2021;13(7):e16649. doi: 10.7759/cureus.16649.
7. Çil ATB, Gümüş S. Solitary Cecal Diverticulitis Mimicking Acute Appendicitis in a Child: Intraoperative Diagnosis. Sisli Etfal Hastan Tip Bul. 2019;53(4):433-436. doi: 10.14744/SEMB.2017.76588.
8. Stollman N, Raskin JB. Diverticular disease of the colon. Lancet. 2004;363(9409):631-639. doi: 10.1016/S0140-6736(04)15597-9.
9. Hildebrand P, Kropp M, Stellmacher F, Roblick UJ, Bruch HP, Schwandner O. Surgery for right-sided colonic diverticulitis: results of a 10-year-observation period. Langenbecks Arch Surg. 2007;392(2):143-147. doi: 10.1007/s00423-006-0109-6.
10. Turner GA, O'Grady MJ, Purcell RV, Frizelle FA. The Epidemiology and Etiology of Right-Sided Colonic Diverticulosis: A Review. Ann Coloproctol. 2021;37(4):196-203. doi: 10.3393/ac.2021.00192.0027.
11. Cave AJ. Appendix Vermiformis Duplex. J Anat. 1936;70(Pt 2):283-292.
12. Connolly D. Solitary caecal diverticulitis. Ulster Med J. 2006;75(2):158.
13. Uhe I, Meyer J, Viviano M, et al. Caecal diverticulitis can be misdiagnosed as acute appendicitis: a systematic review of the literature. Colorectal Dis. 2021;23(10):2515-2526. doi: 10.1111/codi.15818.
14. Wyble EJ, Lee WC. Cecal diverticulitis: changing trends in management. South Med J. 1988;81(3):313-316. doi: 10.1097/00007611-198803000-00006.
15. Abogunrin FA, Arya N, Somerville JE, Refsum S. Solitary caecal diverticulitis--a rare cause of right iliac fossa pain. Ulster Med J. 2005;74(2):132-133.
16. Chou YH, Chiou HJ, Tiu CM, et al. Sonography of acute right side colonic diverticulitis. Am J Surg. 2001;181(2):122-127. doi: 10.1016/s0002-9610(00)00568-7.
17. Brettner A, Euphrat EJ. Radiological significance of primary enterolithiasis. Radiology. 1970;94(2):283-288.
18. Jang HJ, Lim HK, Lee SJ, Lee WJ, Kim EY, Kim SH. Acute diverticulitis of the cecum and ascending colon: the value of thin-section helical CT findings in excluding colonic carcinoma. AJR Am J Roentgenol. 2000;174(5):1397-1402. doi: 10.2214/ajr.174.5.1741397.
19. Lane JS, Sarkar R, Schmit PJ, Chandler CF, Thompson JE Jr. Surgical approach to cecal diverticulitis. J Am Coll Surg. 1999;188(6):629-634; discussion 634-635. doi: 10.1016/s1072-7515(99)00043-5.
20. Wood EH, Sigman MM, Hayden DM. Special Situations in the Management of Diverticular Disease. Clin Colon Rectal Surg. 2021;34(2):121-126. doi: 10.1055/s-0040-1716704.
21. Lee JM, Bai P Chang J, El Hechi M, et al. Hartmann's Procedure vs Primary Anastomosis with Diverting Loop Ileostomy for Acute Diverticulitis: Nationwide Analysis of 2,729 Emergency Surgery Patients. J Am Coll Surg. 2019;229(1):48-55. doi: 10.1016/j.jamcollsurg.2019.03.007.
22. Francis NK, Sylla P, Abou-Khalil M, et al. EAES and SAGES 2018 consensus conference on acute diverticulitis management: evidence-based recommendations for clinical practice. Surg Endosc. 2019;33(9):2726-2741. doi: 10.1007/s00464-019-06882-z.
23. Kaya C, Celayir MF, Bozkurt E, Omeroglu S, Guven O, Mihmanli M. Solitary caecal diverticulitis: Comparison of operative and non operative treatment. J Pak Med Assoc. 2020;70(11):1926-1929. doi: 10.5455/JPMA.12674.