ACTA MEDICA
formerly Hacettepe Medical Journal

www.actamedica.org

Volume 57
Issue 2
June 2026

Online ISSN: 2147-9488

from the seniors to the students

ACTA MEDICA

formerly Hacettepe Medical Journal

Volume **57** • Issue **2** • June **2026**

Online ISSN: 2147-9488

www.actamedica.org

ACTA MEDICA
formerly Hacettepe Medical Journal

Acta Medica

ISSN (Online): 2147-9488

www.actamedica.org

Cilt 57, Sayı 2, Haziran 2026

Sahibi
Hacettepe Üniversitesi
Tıp Fakültesi adına sahibi
Hakan Göker

Sorumlu Yazı İşleri Müdürü
Gözde Yazıcı

Yayının Türü
Yaygın Süreli Yayının

Yayın Sıklığı ve Dili
Üç aylık, İngilizce

Baş Editörler
Gözde Yazıcı
Sevinç Sarınç

Editöryal İletişim
Hacettepe Üniversitesi
Tıp Fakültesi Dekanlığı
06100 Sıhhiye - Ankara, Türkiye
E-posta: editor@actamedica.org

Yayıncı
Hacettepe Üniversitesi
Tıp Fakültesi Dekanlığı
06100 Sıhhiye - Ankara, Türkiye
Telefon: 0 312 305 10 80
Belgeç (faks): 0 312 310 05 80
E-posta: tipmaster@hacettepe.edu.tr

Yayıncılık Hizmetleri
Akdema Bilişim ve Yayıncılık
E-posta: bilgi@akdema.com
Web: www.akdema.com

Acta Medica

ISSN (Online): 2147-9488

www.actamedica.org

Volume 57, Issue 2, June 2026

Owner
Owner on behalf of the Hacettepe University
Faculty of Medicine
Hakan Göker

Administrator
Gözde Yazıcı

Publication Type
Peer-reviewed journal

Publication Frequency and Language
Quarterly, English

Editors-in-Chief
Gözde Yazıcı
Sevinç Sarınç

Editorial Office
Hacettepe University
Faculty of Medicine
06100 Sıhhiye - Ankara, Türkiye
Email: editor@actamedica.org

Publisher
Hacettepe University
Faculty of Medicine
06100 Sıhhiye - Ankara, Türkiye
Phone: 0 312 305 10 80
Fax: 0 312 310 05 80
Email: tipmaster@hacettepe.edu.tr

Publishing Services
Akdema Informatics and Publishing
Email: bilgi@akdema.com
Web: www.akdema.com

ACTA MEDICA
formerly Hacettepe Medical Journal

Editorial Board

ACTA MEDICA
formerly Hacettepe Medical Journal

Contents

Research Articles

Letter to the Editor

Acta Medica 2026; Early View: 91-101 DOI: 10.32552/actamedica.2026.1066

acta medica

RESEARCH ARTICLE

Mismatch Repair (MMR) protein expression in liver tissue and cancer: A study utilizing Human Protein Atlas Data

Meral Üner[1]

[1]Department of Pathology, Faculty of Medicine, Hacettepe University, Ankara, Türkiye

Abstract

Objective: Alterations in mismatch repair (MMR) protein expression have been reported in liver cancers, including hepatocellular carcinoma and cholangiocarcinoma. Loss or reduced expression of MMR proteins is commonly associated with microsatellite instability (MSI), which has important prognostic and therapeutic implications. Immunohistochemistry represents a cost-effective and reliable method for detecting deficient MMR (dMMR). However, due to potential technical pitfalls and variable staining patterns, careful interpretation of MMR immunostaining is required.

Materials and Methods: In this study, virtual slides of tumor microarray cores from 35 patients were re-evaluated. The cases were stained with different antibody clones targeting MLH1, PMS2, MSH2, and MSH6. Based on the percentage of nuclear staining, immunohistochemical expression patterns were categorized as: (1) total expression loss (<25% nuclear staining), (2) focal expression loss (25-75% nuclear staining), and (3) intact expression (>75% nuclear staining).

Results: Among 35 cases (21 hepatocellular carcinomas and 14 cholangiocarcinomas), clone-dependent variability in MMR immunohistochemistry performance was observed. Using the most effective antibody clones, potential dMMR patterns in hepatocellular carcinoma were identified in 25% of cases for MLH1, 50% for MSH2, and 50% for MSH6. In contrast, no dMMR-like pattern was observed in cholangiocarcinoma for any MMR protein. PMS2 (single available clone, CAB010235) demonstrated minimal expression in normal liver tissue and variable staining in tumors, suggesting technical limitations rather than true biological loss. These findings indicate that suboptimal antibody clones may artificially increase the apparent dMMR rate.

Conclusion: This study evaluated MMR protein expression patterns in hepatocellular carcinoma and cholangiocarcinoma using data from the Human Protein Atlas, an open-access resource. Although immunohistochemistry remains a cost-effective and reliable method, factors such as fixation quality, antibody clone sensitivity, and careful evaluation of internal controls must be considered. Appropriate antibody clone selection is therefore essential for accurate interpretation of MMR immunostaining.

Keywords: MMR, liver, immunohistochemistry, cancer, human protein atlas

Introduction

Mismatch repair (MMR) proteins play a critical role in maintaining genomic stability by correcting DNA replication errors [1,2]. Loss of function of at least one of the four key MMR proteins—MLH1, PMS2, MSH2, or MSH6—leads to impairment of the DNA mismatch repair system and results in microsatellite instability (MSI-H), a hypermutable condition observed in various tumor types. Loss of MMR protein expression is therefore

Corresponding author: Meral Üner • **Email:** bugdayci.meral@gmail.com

Received: October 4, 2024 **Accepted:** March 30, 2026 **Published online:** June 28, 2026

commonly associated with MSI-H, a genomic instability phenotype that may arise from either somatic or germline alterations [3].

The frequency of deficient MMR (dMMR) or MSI-H in liver tumors varies according to tumor type and histologic grade [4-7]. In well-differentiated hepatocellular carcinomas (grade 1-2), the frequency of dMMR/MSI-H is generally lower than that observed in poorly differentiated tumors (grade 3), likely reflecting lower proliferative activity and more preserved DNA repair capacity [8].

Several methods, including immunohistochemistry (IHC), polymerase chain reaction (PCR), and next-generation sequencing (NGS), are used to evaluate MMR deficiency in tumors. Among these, immunohistochemistry is a widely used, cost-effective, and practical method for detecting MMR dysfunction [9-11]. However, this technique has several technical limitations and interpretative challenges that require expertise in evaluating staining patterns in both normal and neoplastic tissues. Importantly, MMR protein staining patterns may vary among different tissues, tumor types, and even within the same tumor. In addition, expression of MMR proteins is highly sensitive to tissue fixation conditions [12]. Focal loss of staining may arise from tumor subclones with MMR dysfunction, but it may also result from fixation-related artifacts or abnormal genetic alterations, such as novel mutations or aberrant cytoplasmic accumulation of mutant proteins [13,14].

The present study aims to highlight the technical and interpretative challenges associated with MMR immunohistochemistry, with particular emphasis on antibody clone characteristics, tissue fixation quality, and the role of pathologist expertise. Liver tumors were selected as a model because hepatocellular carcinoma and cholangiocarcinoma display distinct biological and proliferative characteristics. These differences provide an opportunity to demonstrate clone- and tissue-dependent variability in MMR staining patterns and the potential risk of misinterpretation, particularly in well-differentiated, low-proliferation tumors such as hepatocellular carcinoma. Using data derived from the Human Protein Atlas, we evaluated MMR staining patterns across different antibody clones and tissue types in order to emphasize the importance of internal controls and technical awareness for accurate interpretation of MMR immunohistochemistry.

Materials and methods

Study cohort and tissue microarray (tma) core selection

Adult patients with hepatocellular carcinoma or cholangiocarcinoma whose immunohistochemical tissue microarray (TMA) data were available in the open-access Human Protein Atlas (HPA) database were included in the study cohort [15-17].

Immunohistochemical staining for proteins encoded by four MMR genes—MLH1 (clone 1: CAB013294, clone 2: HPA052707, clone 3: CAB070868), PMS2 (clone 1: CAB010235), MSH2 (clone 1: CAB009572, clone 2: CAB070867, clone 3: CAB080131), and MSH6 (clone 1: CAB009091, clone 2: HPA028376, clone 3: HPA028446, clone 4: CAB070870)—was evaluated in the available TMA cores. Normal liver tissue cores were also assessed as controls.

Clinical information including tumor type, age, and sex was obtained from the Human Protein Atlas database (Supplementary Table 1 and Supplementary Table 2) (https://www.proteinatlas.org, accessed August 25 2023).

Evaluation of protein-level expression

Evaluation of MMR protein expression in normal liver tissue cores: In normal liver tissue cores, nuclear staining in hepatocytes, cholangiocytes, fibroblasts, and inflammatory cells was considered the expected staining pattern (0 = loss, 1 = intact) (Figure 1). These cell populations were also used as internal controls for the evaluation of tumor tissues

(Supplementary Table 1 and Supplementary Table 2).

Evaluation of MMR protein expression in neoplastic cores: In tumor cores, peritumoral stromal cells—including hepatocytes, cholangiocytes, fibroblasts, and inflammatory cells—were assessed as internal controls based on their nuclear staining status (0 = loss, 1 = intact).

To distinguish fixation-related staining defects from clone-related staining variability, each core was evaluated using a modified H-score calculated by multiplying the percentage of positive cells by staining

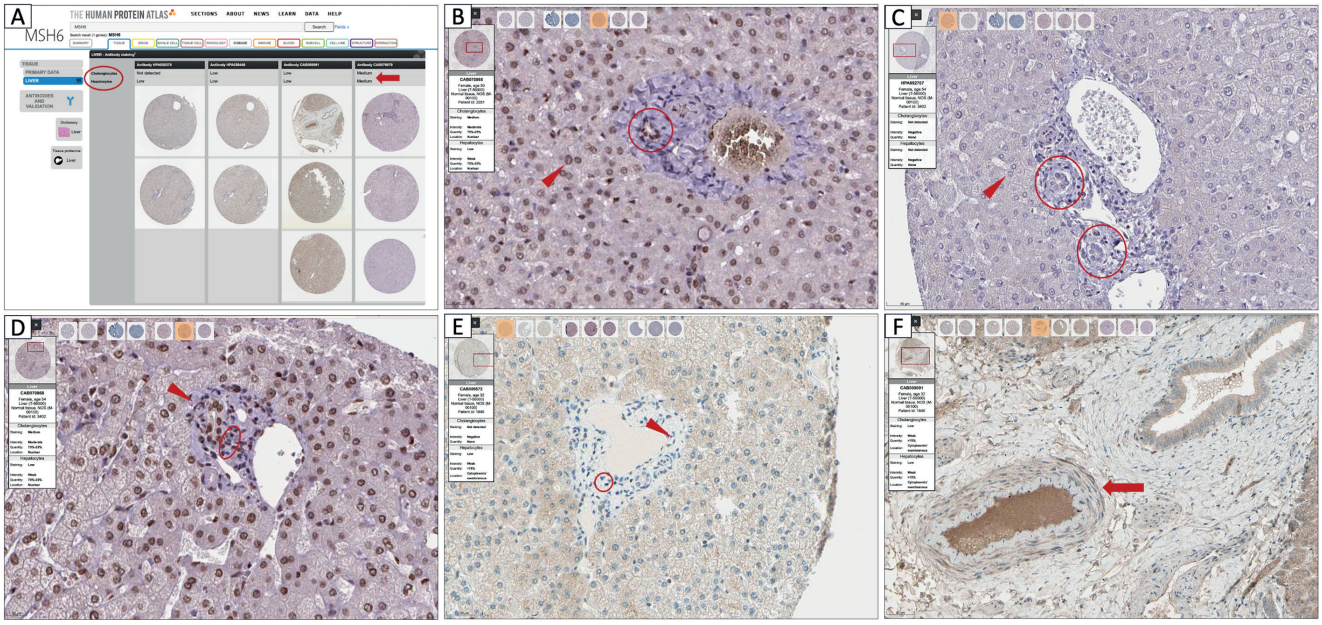


Figure 1. Normal liver tissue expression patterns of MMR antibodies obtained from the Human Protein Atlas (HPA)

Representative virtual tissue microarray (TMA) cores from the Human Protein Atlas (https://www.proteinatlas.org, accessed August 25 2023) showing immunohistochemical staining patterns of mismatch-repair (MMR) proteins in non-neoplastic liver tissue. (A) Overview of MSH6 antibody clones (HPA028376, CAB009091, CAB070870) demonstrating variable nuclear positivity in cholangiocytes and hepatocytes. (B) Intact nuclear expression of MLH1 (clone CAB013294) in hepatocytes (arrowhead) and cholangiocytes (circle) (Patient ID 2251, female, age 50, liver). (C) Total loss of MLH1 nuclear staining in both hepatocytes (arrowhead) and cholangiocytes (circle) (Patient ID 3402, female, age 54, liver). (D) Intact MLH1 nuclear staining in fibroblasts (arrowhead) and scattered inflammatory cells (circle) (Patient ID 3402, female, age 54, liver). (E) Complete loss of MSH2 nuclear staining (clone CAB009572) with non-specific cytoplasmic background in hepatocytes and stromal cells (arrowhead and circle) (Patient ID 1846, female, age 32, liver). (F) Suboptimal MSH6 staining pattern (clone CAB009091) showing absence of expression in vascular endothelial and smooth-muscle cells (arrow), as well as in hepatocytes and cholangiocytes; mild non-specific cytoplasmic background staining is observed in all cell types (Patient ID 1846, female, age 32, liver). Scale bars, section identifiers, and antibody clone codes are displayed in each panel.

intensity (0-3 scale: 0 = negative, 1 = weak, 2 = moderate, 3 = strong) (Supplementary Table 1).

Based on the percentage of tumor cells showing nuclear staining, immunohistochemical expression patterns were categorized as: 1) Total expression loss (<25% nuclear staining), 2) Focal expression loss (25-75% nuclear staining) and 3) Intact expression (>75% nuclear staining). These categories corresponded to the following MMR status groups:

1.dMMR (deficient MMR): Loss of nuclear protein expression of at least one MMR protein (MLH1, PMS2, MSH2, or MSH6) in tumor cells while internal control cells retained intact staining (Figure 2).

2a.fdMMR (focal deficient MMR): Loss of nuclear staining in a subset of tumor cells, reflecting intratumoral heterogeneity, with preserved staining in surrounding internal control cells.

2b.Fixation-related staining loss: Concurrent loss of staining in both tumor cells and surrounding internal control cells, suggesting technical artifact rather than true biological loss.

3.pMMR (proficient MMR): Preserved nuclear staining of all four MMR proteins (MLH1, PMS2, MSH2, and MSH6) in tumor cells (Figure 2).

Results

Demographic findings - All individuals included in the study were adults, including those contributing normal liver tissue used as controls. The mean age of the patients (n=35) was 63.62 years (range: 24-82 years).

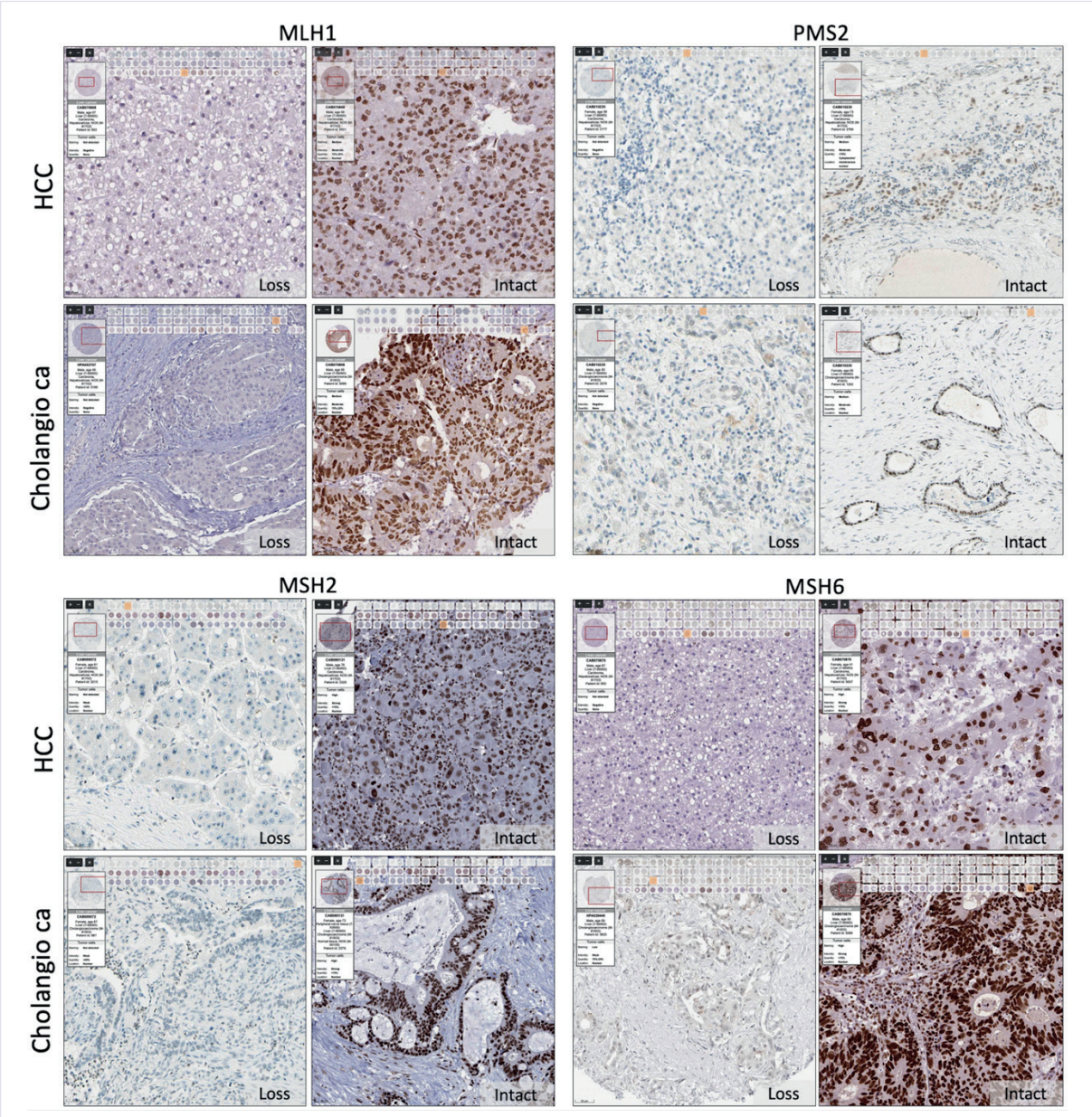


Figure 2. Neoplastic tissue expression patterns of MMR antibodies obtained from the Human Protein Atlas (HPA)

Representative virtual tissue microarray (TMA) cores from the Human Protein Atlas (https://www.proteinatlas.org, accessed August 25 2023) showing immunohistochemical staining patterns of MLH1, PMS2, MSH2, and MSH6 in hepatocellular carcinoma (HCC) and cholangiocarcinoma (Cholangio ca). Panels on the left depict loss of nuclear expression in neoplastic cells, whereas panels on the right show intact nuclear staining. Patient information, antibody clone, and tumor type are indicated in the upper-left box of each image. Each panel also includes the HPA dataset identifier, staining intensity, percentage of positive cells, and localization of staining, which may or may not correspond precisely to the current study's semiquantitative evaluation.

MLH1 – Loss: Patient ID 2767 (M, 58 y, liver hepatocellular carcinoma, clone CAB013294); Intact: Patient ID 3402 (F, 54 y, liver hepatocellular carcinoma, clone CAB013294). PMS2 – Loss: Patient ID 2251 (F, 50 y, liver cholangiocarcinoma, clone CAB010235); Intact: Patient ID 3402 (F, 54 y, liver cholangiocarcinoma, clone CAB010235). MSH2 – Loss: Patient ID 1846 (F, 32 y, liver hepatocellular carcinoma, clone CAB080131); Intact: Patient ID 2251 (F, 50 y, liver cholangiocarcinoma, clone CAB080131). MSH6 – Loss: Patient ID 1846 (F, 32 y, liver hepatocellular carcinoma, clone CAB070870); Intact: Patient ID 2251 (F, 50 y, liver cholangiocarcinoma, clone CAB070870).

HCC = hepatocellular carcinoma; Cholangio ca = cholangiocarcinoma; Loss = absence of nuclear staining in tumor cells; Intact = retained nuclear staining in tumor cells. Scale bars, dataset identifiers, and clone codes are displayed on each panel.

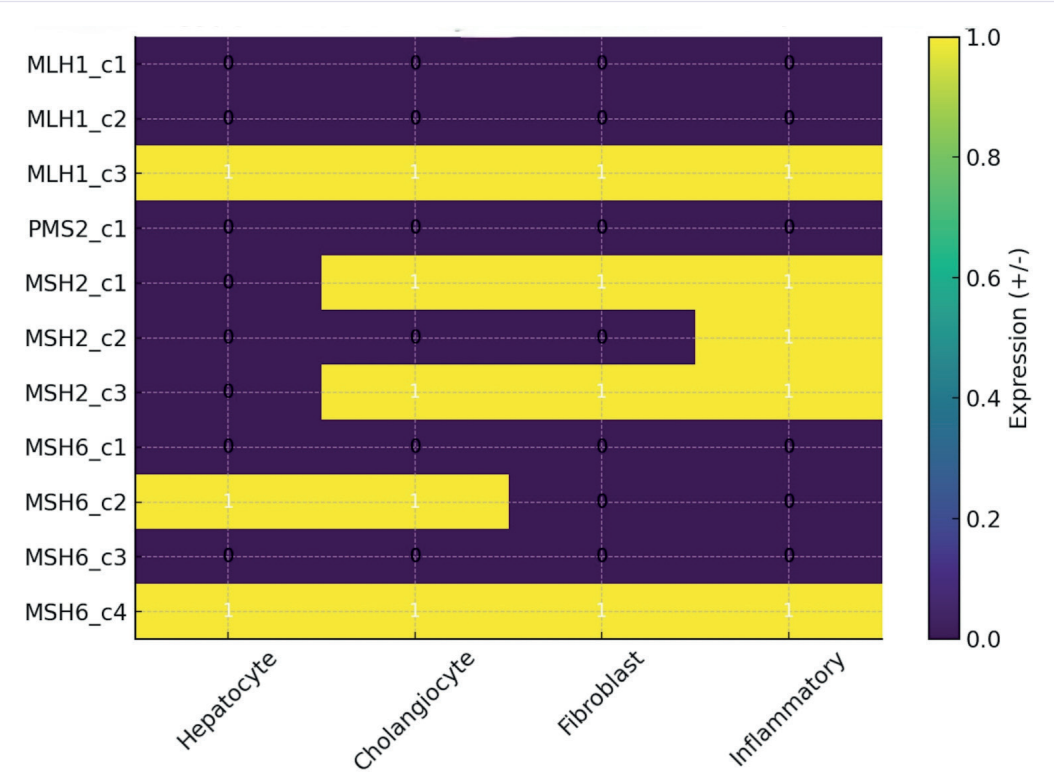


Figure 3. Normal tissue MMR antibody expression pattern by cell type (HPA data)

Heatmap representing the presence (+) or absence (–) of MMR protein expression across different liver cell types based on the Human Protein Atlas dataset. Each cell indicates whether nuclear immunoreactivity was observed in hepatocytes, cholangiocytes, fibroblasts, and inflammatory cells for each antibody clone. Consistent nuclear expression was observed with MLH1 clone 3, MSH2 clone 3, and MSH6 clone 4, while PMS2 clone 1 and early MSH6 clones showed no expression in any cell type. This visualization highlights the variability of baseline expression among clones in non-neoplastic liver tissue, underlining the importance of antibody validation for hepatic immunohistochemistry.

Among the patients, 21 had hepatocellular carcinoma and 14 had cholangiocarcinoma. A summary of the demographic characteristics is provided in Table 1.

Table 1. Patients' demographic data

	HCC	Cholangio ca	Total
Number of patients (n)	21	14	35
Age (year)			
mean	63,2	64,3	63,6
min	24	52	24
max	82	79	82
Gender			
F	8	9	17
M	13	5	18

HCC: Hepatocellular carcinoma, Cholangio ca: Cholangiocarcinoma, F: Female, M: Male.

Clone dependent antibody evaluation results

MLH1 - Based on the evaluation of staining patterns in control tissues, clone 3 was identified as the most effective antibody clone (Figure 3, Supplementary Table 3). In all normal liver tissue cores, hepatocytes, cholangiocytes, fibroblasts, and inflammatory cells showed nuclear positivity for MLH1, indicating intact expression. In neoplastic tissues, both hepatocellular carcinomas and cholangiocarcinomas also demonstrated preserved staining in internal control cells (Figure 4 and Figure 5). Among hepatocellular carcinomas, one tumor showed total expression loss (<25% nuclear staining) and one tumor demonstrated focal expression loss (25-

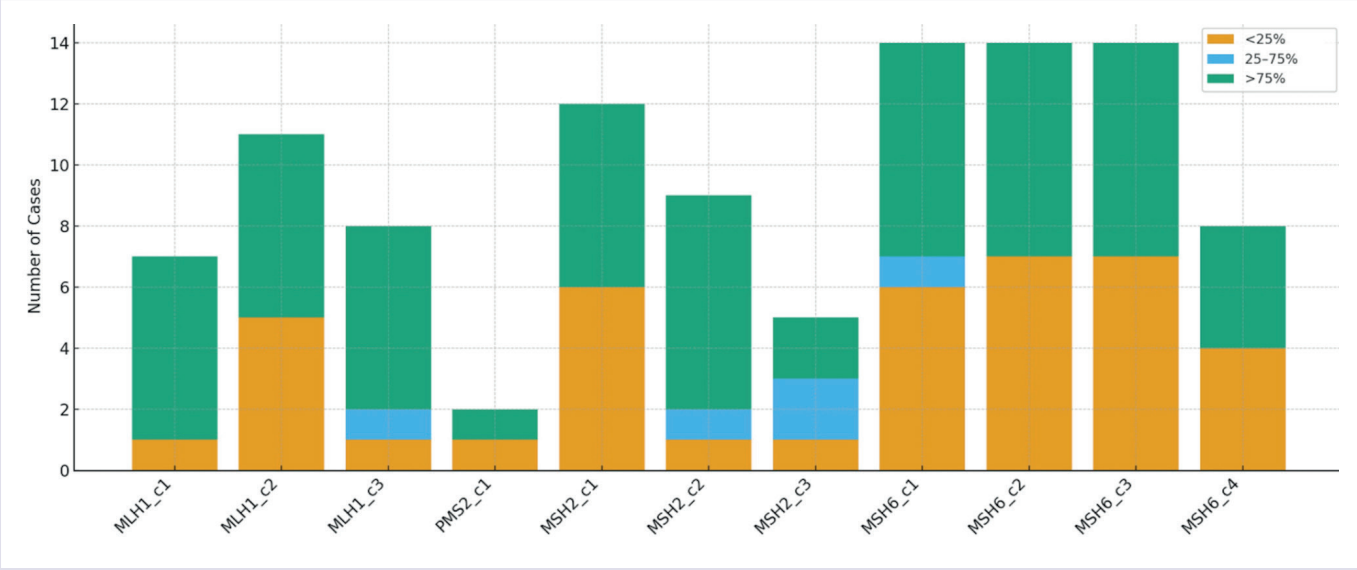


Figure 4. Expression categories across MMR antibody clones in HCC and CCA

Stacked bar chart demonstrating the distribution of MMR IHC expression categories across different antibody clones (MLH1, PMS2, MSH2, MSH6). Expression categories were grouped as <25%, 25–75%, and >75% of tumor cell staining in hepatocellular carcinoma (HCC) and cholangiocarcinoma (CCA) cases. Values represent number of evaluated cases per clone (based on the Human Protein Atlas dataset).

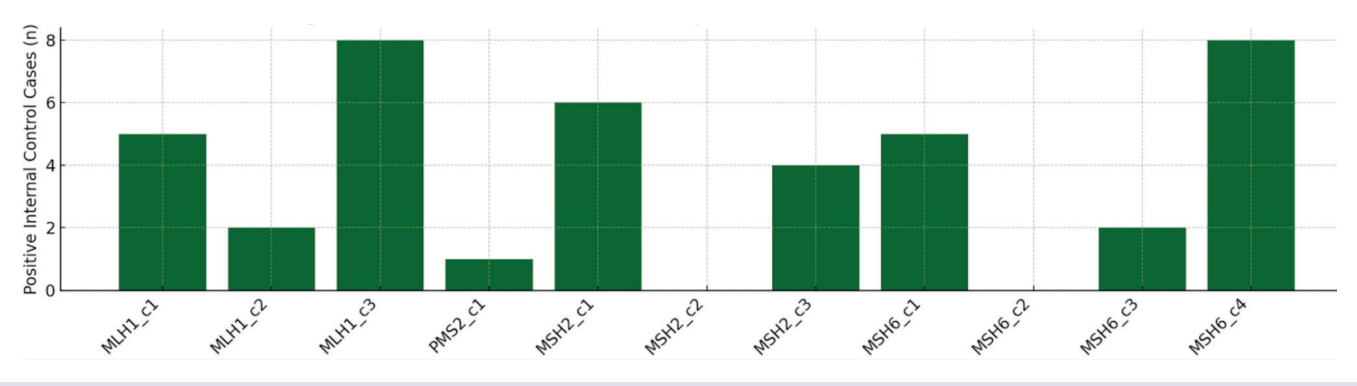


Figure 5. Positive internal control staining across MMR antibody clones

Bar chart showing the number of cases with retained internal control staining (non-neoplastic hepatocytes, cholangiocytes, stromal or inflammatory cells) for each MMR antibody clone. This reflects technical adequacy and clone-specific variability in baseline tissue immunoreactivity (based on the Human Protein Atlas dataset).

75% nuclear staining), whereas six of the eight tumors showed intact expression (>75% nuclear staining). In cholangiocarcinomas, all four tumors showed intact expression. No cases of total or focal expression loss were observed.

PMS2 - Only one antibody clone was available for PMS2 evaluation. This clone demonstrated weak or absent staining in normal liver tissue cores (Figure 3, Supplementary Table 3). None of the hepatocytes, cholangiocytes, fibroblasts, or inflammatory cells showed nuclear staining. In neoplastic cores, four of the six cholangiocarcinoma cores and one of the six hepatocellular carcinoma cores demonstrated preserved staining in internal control cells (Figure 4 and Figure 5). Among hepatocellular carcinomas, five tumors showed total expression loss (<25% nuclear staining) and one tumor showed intact expression (>75% nuclear staining). No cases of focal expression loss were observed. In cholangiocarcinomas, one tumor showed total expression loss (<25% nuclear staining), while five tumors demonstrated intact

expression (>75% nuclear staining). No focal expression loss was detected (25%- 75%).

MSH2 - Based on control tissue staining patterns, antibody clone 3 was identified as the most effective clone (Figure 3). In normal liver tissue cores, hepatocytes showed no nuclear staining, whereas cholangiocytes, fibroblasts, and inflammatory cells demonstrated nuclear positivity for MSH2. In neoplastic tissues, both hepatocellular carcinomas and cholangiocarcinomas showed preserved staining in internal control cells (Figure 4 and Figure 5). Among hepatocellular carcinomas, two tumors demonstrated focal expression loss (25-75% nuclear staining), while two tumors showed intact expression (>75% nuclear staining). No cases of total expression loss were observed. In cholangiocarcinomas, all seven tumors demonstrated intact expression. No cases of total or focal expression loss were detected.

MSH6 - Evaluation of staining patterns in control tissues indicated that clone 4 was the most effective antibody clone (Figure 3). In normal liver tissue cores, hepatocytes, cholangiocytes, fibroblasts, and inflammatory cells all demonstrated nuclear positivity for MSH6, indicating intact expression. In neoplastic tissues, both hepatocellular carcinomas and cholangiocarcinomas showed preserved staining in internal control cells (Figure 4 and Figure 5).

Among hepatocellular carcinomas, four tumors demonstrated total expression loss (<25% nuclear staining), while four tumors showed intact expression (>75% nuclear staining). No cases of focal expression loss were observed. Among cholangiocarcinomas, all four tumors showed intact expression. No cases of total or focal expression loss were detected.

Most effective antibody clone-dependent MMR evaluation results

MMR status was further evaluated using the most effective antibody clones identified for each MMR protein. When total or focal expression loss was considered indicative of possible dMMR, the following results were observed:

For hepatocellular carcinoma, possible dMMR rates were 25% for MLH1, 50% for MSH2, and 50% for MSH6.

For cholangiocarcinoma, no cases of possible dMMR were observed for MLH1, MSH2, or MSH6.

When antibody clones other than the most effective clone were used (representing suboptimal staining conditions), the apparent frequency of possible dMMR increased due to ineffective staining (Figure 4 and Figure 5, Supplementary Table 4).

Under these conditions, possible dMMR rates ranged from 25-100% for MLH1, 100% for MSH2, and 85-100% for MSH6 in hepatocellular carcinoma.

For cholangiocarcinoma, possible dMMR rates ranged from 20-100% for MLH1, 66-75% for MSH2, and 80-100% for MSH6.

For PMS2, only one antibody clone was available, which showed suboptimal staining. Using this clone, the apparent dMMR rate was 83% for hepatocellular carcinoma and 16% for cholangiocarcinoma.

Statistical evaluation of MMR expression patterns

Based on the available Human Protein Atlas (HPA) dataset, both categorical and continuous variables were analyzed using appropriate non-parametric statistical tests.

The proportions of loss versus intact expression were compared between hepatocellular carcinoma and cholangiocarcinoma using Fisher's exact test, whereas differences in staining intensity (0-3 scale) were assessed using the Mann-Whitney U test.

Among the MMR proteins analyzed, MSH2 demonstrated a statistically significant difference between hepatocellular carcinoma and cholangiocarcinoma (Fisher's exact test, $p = 0.015$), indicating a higher frequency of expression loss in cholangiocarcinoma.

A secondary comparison of staining intensity also showed statistical significance for MSH2 ($p = 0.007$) and borderline significance for PMS2 ($p = 0.045$). No other antibodies demonstrated statistically significant differences (all $p > 0.05$).

Given the limited sample size and the observational nature of the HPA dataset, these findings should be

interpreted as descriptive and exploratory rather than confirmatory (Figure 3, Figure 4 and Figure 5).

In supplementary analyses, comparisons of MLH1, PMS2, MSH2, and MSH6 expression levels revealed no statistically significant differences in modified H-score values among the four MMR proteins (Kruskal-Wallis test, $p = 0.407$), suggesting similar overall expression patterns across these markers (Supplementary Table 1).

Binary expression assessment in normal liver tissue demonstrated consistent hepatocellular and stromal staining for MLH1 and MSH6, whereas PMS2 showed minimal expression in normal liver cores (Supplementary Table 2).

Discussion

Hepatocellular carcinoma is the most common primary malignancy of the liver, followed by cholangiocarcinoma as the second most frequent type [18,19]. In this study, we focused on these two major liver tumor types and analyzed the expression status of MMR proteins using virtual slide data available in the Human Protein Atlas (HPA). This open-access platform provides a valuable resource for both pathologists and researchers by enabling systematic evaluation of immunohistochemical staining patterns across different tissues and tumor types.

Mismatch repair deficiency in hepatocellular carcinoma has been reported at varying frequencies in the literature. Bonneville et al. reported a prevalence of 0.8%, whereas Mukai et al., Kawaoka et al., and Cortes et al. reported frequencies of 2%, 2.4%, and 2.9%, respectively [4-6,20]. In the current study, higher apparent dMMR rates (25-50%) were observed compared with previously reported data. This discrepancy is most likely attributable to the limited sample size, which may not accurately represent the overall population.

For cholangiocarcinoma, MMR deficiency has been reported in 4.1% to 10% of intrahepatic cholangiocarcinomas in studies by Ju et al. and Saeed et al. [7,21]. In contrast, no dMMR cases were detected among the cholangiocarcinomas analyzed in the present study, which may similarly be explained by the limited sample size.

From an immunohistochemical perspective, MMR protein expression has been reported to be higher in normal tissues with increased proliferative activity [8]. Similarly, higher MMR expression has been observed in high-grade tumors, which typically demonstrate elevated proliferation rates [8,15]. This association between MMR expression and tumor proliferative activity may partly explain our observation of more frequent expression loss in hepatocellular carcinomas compared with cholangiocarcinomas, as the hepatocellular carcinoma cases included in the present study were predominantly low-grade tumors. This factor may also contribute to the discrepancy between the dMMR frequencies observed in this study and those reported in the literature. In addition to the limited sample size, several methodological factors may explain the higher apparent dMMR rate observed in this study. The Human Protein Atlas database does not represent a true clinical cohort and may therefore be subject to selection bias. Furthermore, MMR immunohistochemistry is highly dependent on antibody clone performance and tissue fixation quality, both of which may influence staining intensity and potentially lead to misinterpretation of reduced staining as true biological loss, particularly in well-differentiated and low-proliferation liver tumors. The use of a single tissue core per tumor may also underrepresent areas with preserved staining. Taken together, the observed dMMR frequency in this study likely reflects technical and sampling variability rather than true biological prevalence. Larger, clinicopathologically well-characterized cohorts with optimized immunohistochemical protocols and external validation are required to more accurately determine the incidence of dMMR in liver cancers.

Assessment of MMR protein expression by immunohistochemistry may present several challenges for pathologists, particularly with respect to pre-analytical variables such as tissue fixation [22]. Normal cells present within the section can serve as internal controls to help determine whether apparent loss of expression reflects fixation-related artifacts or true tumor heterogeneity [22]. However, in cases lacking an appropriate internal control, distinguishing between these possibilities becomes difficult.

Additionally, because each tumor in the HPA dataset was represented by only one or two tissue cores, the impact of intratumoral heterogeneity—particularly focal

or patchy reduction of MMR protein expression—cannot be fully evaluated. In routine diagnostic practice, focal MMR alterations have been documented in several solid tumors, and sampling only one or two tumor areas may therefore underestimate or misclassify MMR status. Furthermore, in routine diagnostic pathology, the selection of tissue blocks and the choice of core sampling sites on H&E sections have a significant influence on interpretation. Preferential sampling of well-fixed, viable, and necrosis-free tumor areas improves the reliability of MMR immunohistochemistry and reduces the likelihood of artifactual staining patterns that may mimic true biological loss. Future studies incorporating multiple-core sampling and whole-slide evaluation, together with careful selection of well-preserved tumor regions on H&E sections, would better address spatial heterogeneity and help validate these observations in a more comprehensive manner.

Another important factor influencing immunohistochemical evaluation is the sensitivity of antibody clones directed against the same antigen. Differences in clone performance may account for the observed variability in staining patterns, and the use of suboptimal antibody clones can further complicate interpretation of MMR immunohistochemistry [23]. In the present study, this effect was clearly demonstrated by the significant staining differences observed among the MLH1, MSH2, and MSH6 antibody clones [23].

The number of liver cancer cases available in the Human Protein Atlas dataset was limited (n = 35), which restricts the statistical power of the analysis and precludes definitive conclusions regarding the true prevalence of MMR deficiency in hepatocellular carcinoma and cholangiocarcinoma. Therefore, the findings of this study should be interpreted as descriptive and exploratory rather than representative of the general population. In addition, each tumor was represented by a single tissue core, which may not adequately capture intratumoral heterogeneity or focal patterns of MMR protein loss. Furthermore, fixation quality and other pre-analytical conditions were not fully documented in the dataset, potentially affecting staining performance and interpretation. Future large-scale, well-annotated cohort studies incorporating comprehensive clinicopathological and molecular correlations are required to validate these observations and to better define the true frequency and clinical significance of MMR deficiency in liver malignancies.

From a clinical perspective, although the prevalence of dMMR/MSI-H in hepatocellular carcinoma and cholangiocarcinoma is low, determination of MMR status may still have important therapeutic implications. Tumors with MSI-H/dMMR across multiple organ systems are eligible for immune checkpoint inhibitor therapy, such as pembrolizumab, based on tumor-agnostic regulatory approval. Therefore, accurate interpretation of MMR immunohistochemistry is essential to avoid false-negative or false-positive assessments and to ensure appropriate patient selection for potential immunotherapy. Although routine MMR testing is not currently recommended in liver cancers because of the low frequency of dMMR, selective testing may be considered in specific clinical or pathological contexts, including young patients, unusual tumor morphology, or a strong personal or family history suggestive of MMR deficiency. Continued refinement of immunohistochemical interpretation standards and increased awareness of potential technical pitfalls will support the accurate identification of rare but clinically actionable dMMR cases.

In conclusion, this study highlights the technical complexity and interpretative challenges associated with MMR immunohistochemistry, particularly in liver pathology. Using hepatocellular carcinoma and cholangiocarcinoma as representative tumor models, we demonstrated that the interpretation of MMR staining is strongly influenced by antibody clone characteristics, tissue fixation quality, and the intrinsic proliferative and morphological features of the tumor. Well-differentiated, low-proliferative tumors such as hepatocellular carcinoma may show weak or heterogeneous staining patterns that complicate interpretation, whereas cholangiocarcinoma, with relatively higher proliferative activity, may provide clearer internal reference staining. These findings emphasize that although immunohistochemistry remains a cost-effective and widely used method for detecting dMMR, careful technical and interpretative evaluation is essential. Awareness of clone-specific staining behavior, fixation-related artifacts, and the appropriate use of internal controls is critical for accurate interpretation. Careful antibody clone selection, optimized technical protocols, and expert pathological evaluation are therefore necessary to avoid misclassification, particularly in tissues with variable fixation characteristics and low proliferative indices such as liver tumors.

Institutional review board statement

Not applicable. All data used in this study are from public databases as described in the methods, for which no ethical approval is required.

Author contribution

Conception and design: M.Ü.; Data acquisition: M.Ü.; Data analysis: M.Ü.; Data interpretation: M.Ü.; Drafting of the manuscript: M.Ü.; Critical revision of the manuscript: M.Ü. The author reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of this study.

Ethical approval

Ethics committee approval and informed consent were not required for this study.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The author declares that this study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The authors declare that this study received no funding.

Generative AI statement

The author declares that no generative AI or AI-assisted technologies were used in the writing or preparation of this study.

References

[1] Prolla TA, Pang Q, Alani E, Kolodner RD, Liskay RM. MLH1, PMS1, and MSH2 interactions during the initiation of DNA mismatch repair in yeast. Science 1994;265(5175):1091-3. **[Crossref]**

[2] Warmerdam DO, Kanaar R. Dealing with DNA damage: relationships between checkpoint and repair pathways. Mutat Res 2010;704(1-3):2-11. **[Crossref]**

[3] Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology 1999;116(6):1453-6. **[Crossref]**

[4] Bonneville R, Krook MA, Kautto EA, et al. Landscape of microsatellite instability across 39 cancer types. JCO Precis Oncol 2017;2017:1-15. **[Crossref]**

[5] Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. Nat Commun 2017;8:15180. **[Crossref]**

[6] Kawaoka T, Ando Y, Yamauchi M, et al. Incidence of microsatellite instability-high hepatocellular carcinoma among Japanese patients and response to pembrolizumab. Hepatol Res 2020;50(7):885-888. **[Crossref]**

[7] Ju JY, Dibbern ME, Mahadevan MS, Fan J, Kunk PR, Stelow EB. Mismatch repair protein deficiency/microsatellite instability is rare in cholangiocarcinomas and associated with distinctive morphologies. Am J Clin Pathol 2020;153(5):598-604. **[Crossref]**

[8] Oshi M, Kim TH, Tokumaru Y, et al. Enhanced DNA repair pathway is associated with cell proliferation and worse survival in Hepatocellular Carcinoma (HCC). Cancers (Basel) 2021;13(2):323. **[Crossref]**

[9] Lindor NM, Burgart LJ, Leontovich O, et al. Immunohistochemistry versus microsatellite instability testing in phenotyping colorectal tumors. J Clin Oncol 2002;20(4):1043-8. **[Crossref]**

[10] Zhang X, Li J. Era of universal testing of microsatellite instability in colorectal cancer. World J Gastrointest Oncol 2013;5(2):12-9. **[Crossref]**

[11] Batte BAL, Bruegl AS, Daniels MS, et al. Consequences of universal MSI/IHC in screening ENDOMETRIAL cancer patients for Lynch syndrome. Gynecol Oncol 2014;134(2):319-25. **[Crossref]**

[12] Fadhil W, Ilyas M. Immunostaining for mismatch repair (MMR) protein expression in colorectal cancer is better and easier to interpret when performed on diagnostic biopsies. Histopathology 2012;60(4):653-5. **[Crossref]**

[13] Shia J, Holck S, Depetris G, Greenson JK, Klimstra DS. Lynch syndrome-associated neoplasms: a discussion on histopathology and immunohistochemistry. Fam Cancer 2013;12(2):241-60. **[Crossref]**

[14] Jaffrelot M, Farés N, Brunac AC, et al. An unusual phenotype occurs in 15% of mismatch repair-deficient tumors and is associated with non-colorectal cancers and genetic syndromes. Mod Pathol 2022;35(3):427-37. **[Crossref]**

[15] The Human Protein Atlas. Available at: https://www.proteinatlas.org (Accessed on August 25, 2023).

[16] Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. Science 2015;347(6220):1260419. **[Crossref]**

[17] Pontén F, Jirström K, Uhlen M. The Human Protein Atlas-a tool for pathology. J Pathol 2008;216(4):387-93. **[Crossref]**

[18] Pellino A, Loupakis F, Cadamuro M, et al. Precision medicine in cholangiocarcinoma. Transl Gastroenterol Hepatol 2018;3:40. **[Crossref]**

[19] Ilyas SI, Gores GJ. Pathogenesis, diagnosis, and management of cholangiocarcinoma. Gastroenterology 2013;145(6):1215-29. **[Crossref]**

[20] Mukai S, Kanzaki H, Ogasawara S, et al. Exploring microsatellite instability in patients with advanced hepatocellular carcinoma and its tumor microenvironment. JGH Open 2021;5(11):1266-1274. **[Crossref]**

[21] Saeed A, Park R, Al-Jumayli M, Al-Rajabi R, Sun W. Biologics, immunotherapy, and future directions in the treatment of advanced cholangiocarcinoma. Clin Colorectal Cancer 2019;18(2):81-90. **[Crossref]**

[22] Torlakovic EE, Nielsen S, Francis G, et al. Standardization of positive controls in diagnostic immunohistochemistry: recommendations from the International Ad Hoc Expert Committee. Appl Immunohistochem Mol Morphol 2015;23(1):1-18. **[Crossref]**

[23] NordiQC. MMR proteins. Available at: https://www.nordiqc.org/epitope.php?id=80

acta medica

RESEARCH ARTICLE

Survival analysis of trauma implants in orthopedic surgery

Melih Oral[1], Ulaş Can Kolaç[1], Saygın Kamacı[1]

[1]Department of Orthopedics and Traumatology, Faculty of Medicine, Hacettepe University, Ankara, Türkiye

Abstract

Objective: This study aimed to evaluate the survivability and success rate of medical trauma implants used in orthopedic trauma surgery. Fracture union rates, postoperative complications, and implant failures were analyzed to determine the overall performance of the implants.

Materials and Methods: The study included 455 patients (621 fractures) who underwent surgical fixation with Oltho Medical orthopedic trauma implants at Hacettepe University between January 2020 and July 2022. Patients with at least six months of follow-up and adequate radiographic evaluation were included. Fracture healing was assessed using the Radiographic Union Scale for Tibia (RUST) and a similar four-cortex scoring system for other fractures. Implant failures, postoperative complications, and patient demographics were recorded.

Results: The mean patient age was 34.2 ± 24.3 years (range: 6–93), with a mean follow-up of 16 months. The overall fracture union rate was 97.2% at six months, with a mean RUST score of 10.5. The implant survival rate was 98.7%, with implant failure observed in six (1.3%) cases. Postoperative complications occurred in 7.3% of patients, including wound complications (6.2%), nonunion (1.3%), and implant failures (1.3%). Although implant-specific survival rates varied among anatomical regions, all fractures ultimately achieved union following appropriate management.

Conclusion: Medical trauma implants demonstrated high survivability and effectiveness in fracture fixation, with a low failure rate comparable to existing literature. Despite some complications, all were successfully managed, supporting the reliability of these implants in orthopedic trauma surgery.

Keywords: orthopedic implants, fracture healing, trauma surgery, implant failure, survivability

Introduction

Bone is a dynamic tissue, and the damage that occurs after trauma initiates the healing process due to the high remodeling and regeneration capacity of the bone [1]. Fracture healing is a physiological process that is affected by biological and biomechanical factors [1,2]. Fracture stabilization and fixation are essential for bone healing, making orthopedic implants indispensable in trauma surgery [3,4]. Successful fracture healing depends on biological, mechanical, and biomechanical factors, as well as appropriate surgical management [5,6]. The path to be followed in the detection and treatment of the fracture is affected by factors such as trauma severity, fracture location, soft tissue condition, patient's medical condition, bone quality, and surgical experience [7].

Fracture healing is a dynamic process influenced by multiple factors [3,8]. Bone healing occurs through two primary mechanisms: primary (direct) healing and secondary (indirect) healing. Primary (direct) healing

Corresponding author: Melih Oral • **Email:** melihoral12@gmail.com

Received: February 18, 2025 **Accepted:** April 23, 2026 **Published online:** June 28, 2026

is rare and occurs when absolute stability is achieved, with interfragmentary movement of less than 0.01 mm and strain below 2% [9]. Secondary (indirect) healing is the most seen type, and it is a form of recovery after fracture end movement and weight bearing. It is seen in all non-surgical treatments, also in intramedullary nailing, external fixation, and some internal fixation treatments in complicated comminuted fractures [10,11].

The design of orthopedic implants has shown certain changes from the time it was started to use [12]. Modern implants improve bone healing through improved biocompatibility, mechanical stability, and surface chemistry [12]. Implant-bone healing mechanism differs depending on implant surface chemistry and topography [13]. Peri-implant incompatibility causes complications as nonunion and multiple failure due to implant insufficiency [14,15]. Nonunion occurs in approximately 2% of all fractures but may reach 20% in diaphyseal fractures [15]. The U.S. Food and Drug Administration (FDA) defines nonunion as the absence of healing within nine months or lack of progression over three consecutive months [16].

Fracture healing is assessed through both radiological and clinical evaluation [17]. While pain during weight bearing is considered clinically significant, radiologically progressive improvement in serial radiographs and cortical bridging are also to be considered [18]. Healing scores in four cortices on anterior-posterior and lateral radiographs has been modified recently, taking into account the bony bridging and non-bridging callus tissue, that has been applied in certain fracture types [19,20].

This study aimed to assess the survivability and success rate of medical trauma implants in various surgical fixation methods across different fracture types. Additionally, we aimed to analyze patient demographics, comorbidities, fracture union rates, and associated complications to further evaluate implant performance.

Material and methods

This study was approved by the Institutional Ethics Committee (IRB No: GO 22/942). We retrospectively analyzed all patients who underwent surgical fixation with Oltho Medical orthopedic trauma implants (Ankara, Turkey) at Hacettepe University between January 2020 and July 2022. Routine postoperative follow-up was conducted at 2 weeks, 6 weeks, 3 months, 6 months, 1 year, and annually thereafter. The study included patients with at least 6 months of follow-up and proper antero-posterior and lateral views of the fracture site. Patients with less than 6 months of follow-up and inadequate radiographs were excluded from the study. A total of 601 patients were identified. Of these, 92 were excluded due to inadequate radiographs, and 54 were excluded for having less than 6 months of follow-up. Immediate postoperative, 3-month, and 6-month radiographs were analyzed for fracture union and implant failure (loosening, breakage, and subsidence). Fracture union was determined as viewing bridging callus formation on radiographs (3 out of 4 cortices in anteroposterior and lateral views if possible] and non-tender/painless palpation of the fracture site). Patient demographic data and clinical notes were retrospectively collected from medical records. Radiological evaluation was conducted independently by a fellowship-trained orthopedic trauma surgeon using the Picture Archiving and Communication System (PACS).

Tibial fracture healing was assessed using the Radiographic Union Scale for Tibia (RUST) score (Table 1) [20,21]. For other fracture types amenable to standard two-view radiographic analysis, a RUST-like scoring system was used. Each of the 4 cortex (anterior, posterior, medial and lateral) was assessed and received a score between 1 to 4 according to the healing stage. A cortex with a visible fracture line without callus formation received 1 point, cortex with a visible fracture line with callus formation received 2 points, cortex with a visible fracture line with bridging callus formation received 3 points and a remodeled cortex without a visible fracture line received 4 points. Then points for the all four cortices were summed up and evaluated. A total of four points were determined as absolute nonunion and 12 points were determined as complete union [20,21].

Table 1. RUST scoring system

Radiologic criteria		
Score per cortex	**Callus**	**Fracture line**
1	Absent	Visible
2	Present	Visible
3	Present	Invisible
4	Remodelled	Invisible

Patient demographics, medical comorbidities, concomitant fractures, fracture type, affected bone, and intraoperative and postoperative complications were recorded. All patients received appropriate postoperative physical therapy, analgesics, anticoagulation, and immobilization based on individual risk factors. The patients were checked for soft tissue problems like wound complications, implant prominence, surgical site infections in every follow-up. Appropriate antibiotic therapy was administered incase of surgical site infection. Fracture non-union was assessed as failure to demonstrate bridging callus and RUST scores of 4 at 6th months follow up. Fracture non-union was addressed with revision fixation surgery with or without bone grafting, implant prominence was addressed with implant removal surgery in at appropriate time following fracture union. Pediatric implants were routinely removed at least 1 year after the surgery.

Statistical analysis

The statistical analysis was performed wit SPSS 22.0 (IBM, Sanfracisco). The descriptive characteristics of the study patients were determined using frequency and percentage, analysis of the scale was performed with mean values and standard deviation. The relationship between patient age and RUST scores was calculated via Pearson's correlation analysis. The differentiation of the score levels according to the descriptive specifics was calculated via T-test, one way variance analysis (ANOVA) and Post-hoc analysis.

Results

The study included 455 patients with 621 fractures Figure 1. The mean age for the study patients was 34.2 ± 24.3 (range; 6 – 93) years. 58% of the patients were male and 42% of the patients were female. The mean follow-up was 16 [range; 9 – 24] months. Associated medical comorbidities were noted in 34.5% of the patients. According to radiological findings, physical examinations and RUST scores the overall fracture union rate was 97,2% at 6th months follow-up. The RUST scores at the 6th-month follow-up was 10.5 which showed union of the fractures. The evaluation of RUST scrores during follow-ups were demonstrated in Figure 2. The overall complication rate was 7,3% following fracture fixation surgery.

The most common location for surgically treated fractures was ankle fractures (99 fractures) followed by forearm fractures (80 fractures) and hip fractures (52 fractures). The anatomic distribution of the fractures were demostrated in Table 2.

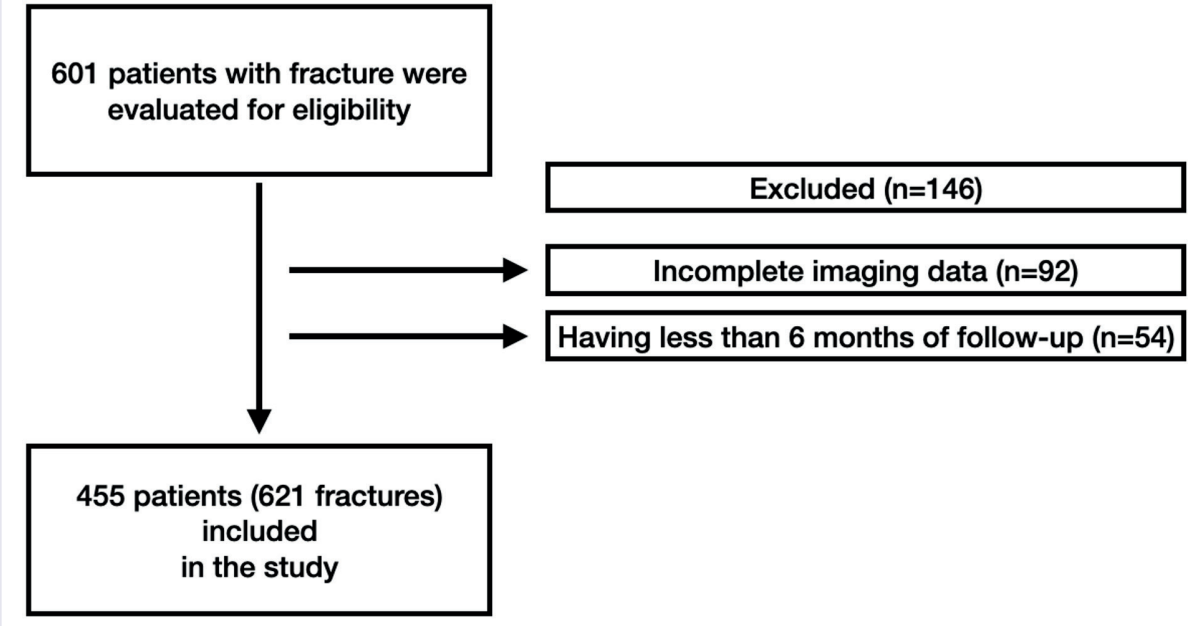


Figure 1. Exclusion criteria for patient eligibility

Postoperative complications were seen in 7.3% of the patients, while 92.7% of the patients were free of complications. In the postoperative course 6.2% of the patients suffered postoperative wound problems that were treated with antibiotics and surgical debridement when needed. Nonunion was seen in 1.3% [6] of the patients who were treated with revision surgery. Implant failure occurred in 1.3% [6] patients which caused

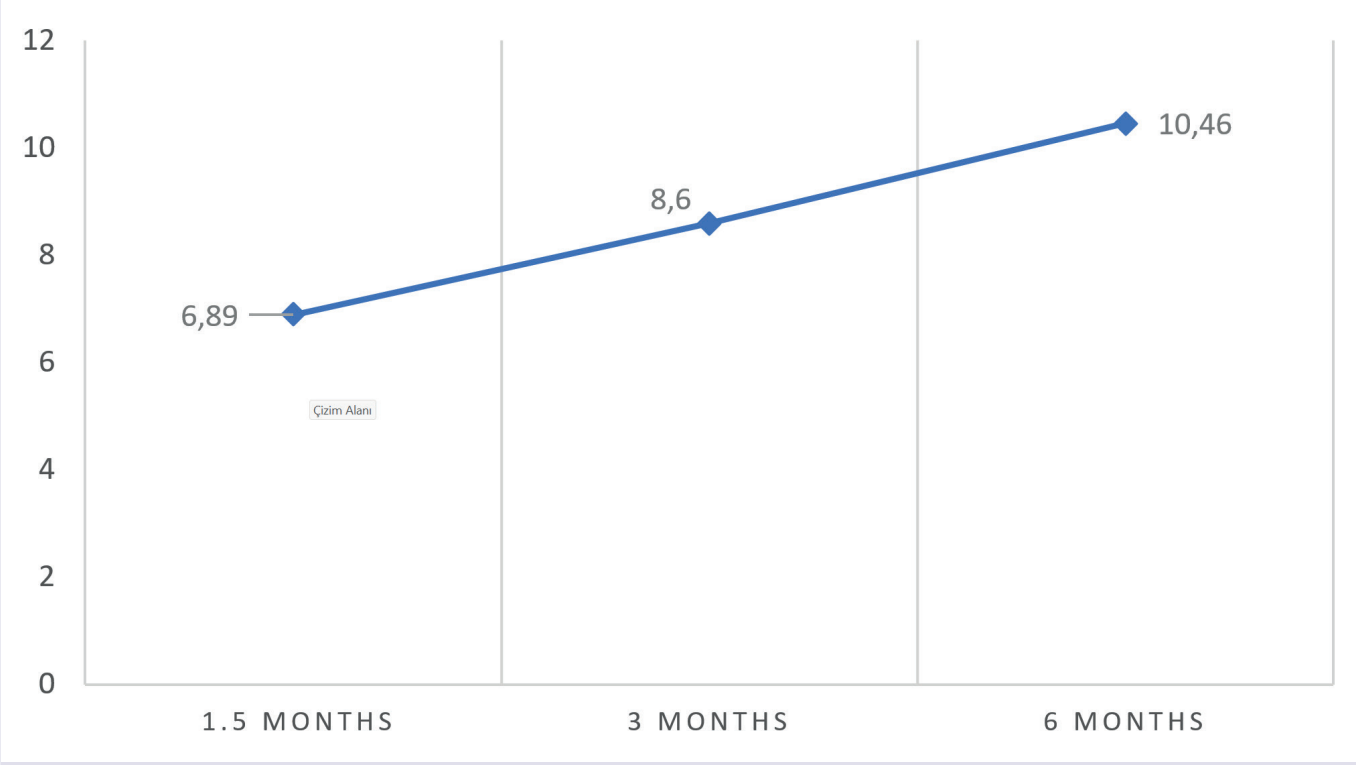


Figure 2. The evaluation of RUST scores during follow-ups

Table 2. The distrubition of fracture types

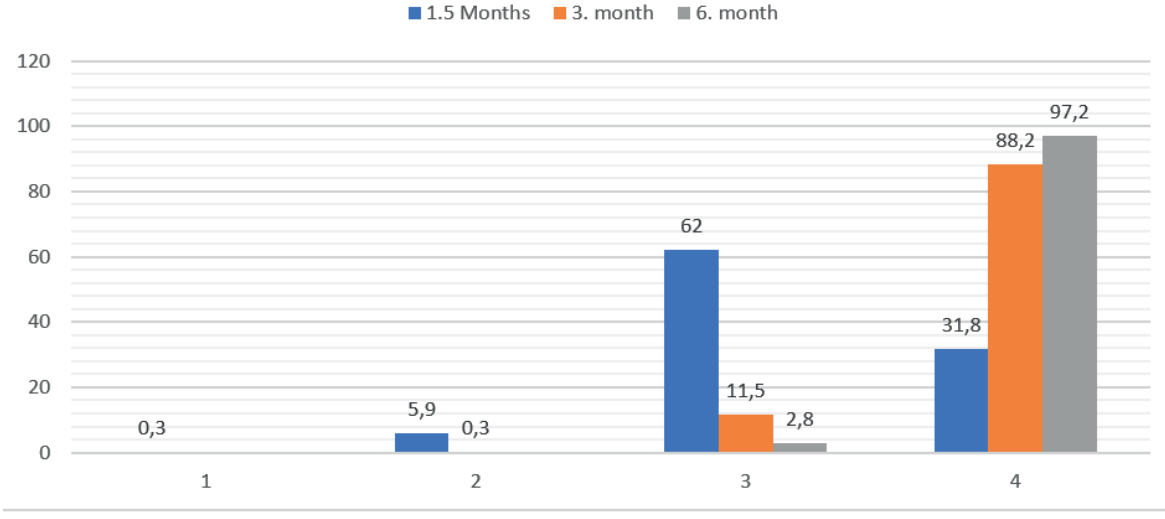
Fracture types	Number of fractures	Fracture types	Number of fractures
Proximal femur fracture	52	Tibia plateau fracture	27
Clavicula fracture	44	Trimalleolar fracture	24
Distal humerus fracture	43	Proximal ulna fracture	20
Lateral malleolar + Bimalleolar fracture	42	Femur Shaft fracture	19
Tibia shaft fracture	35	Metatarsal fracture	18
Distal radius fracture	34	Metacarpal fracture	15
Calcaneus fracture	34	Ulna shaft fracture	13
Proximal humerus fracture	33	Radial shaft fracture	13
Distal femur fracture	33	Radial head fracture	12
Humerus shaft fracture	32	Talus fracture	12
Distal tibia fracture	30	Phalanx fracture	8
Both bone forearm fracture	28		
TOTAL		621	

revision fixation procedure. One patient with both bone forearm fracture suffered radial nerve palsy that was solved at the 3rd month follow up.

The fracture union rate was evaluated via RUST scores which significantly improved during follow-ups, Figure 3 . At the 6th weeks follow up, 62% of the fractures showed 3 and 31,8% showed 4 cortex callus formation. At 6th months follow up 97.2% of the fractures showed 4 cortices callus formation that improved significantly ($p< 0.05$).

A total of 996 different implants were used during the surgical treatment course. The distribution of the implants used during the surgical treatment were demonstrated in Table 3.

The most important aspect of orthopedic trauma implants is to bear biological loads without breakage till fracture union. Since the clinical outcomes depends on mutlifactorial viariables like fracture type, location, patient related ractors like comorbidities, treatment compatibility , and surgical factors like techniqe and proper implant usage, we decided to analyze implant survival till union and operative complications. Below are the individual analysis of specific implants.



Measurement	Mean	SD	F	p	Bonferroni
Score at 1.5 months	6,890	1,131	1389,638	**0,000**	**2>1; 3>1,2**
Score at 3. month	8,600	1,334			
Score at 6. month	10,460	1,687			

Figure 3. RUST scores and cortex ratio percentile according to follow ups

Table 3. The distrubition of implants used in the study

Plates	Quantity	Plates	Quantity
Mini Plate	36	Olecranon Plate	25
3.5 LC-LCP compression Plate	35	LCP Superior Anterior Clavicle Plate, Locked Lateral Extension	24
3.5mm Locking Reconstruction Plate	35	Pediatric Plate 3.5 mm	22
Distal Radius Plate Anatomic	35	LCP Proximal Medial Tibia Plate	21
1/3 Tubular Plate	34	Pediatric Plate 4.5 mm	19
Distal Lateral Femur Plate + Minimal Invasive	33	4.5 mm Reconstruction Locking Plate	18
Distal Fibula 3.5mm Plate Anatomic	32	T-Buttres Plate	17
Distal Medial Humerus Plate	32	Distal Lateral Dorsal Humerus Plate	17
Distal Medial Tibia Plate	31	LCP Clavicle Hook Plate	14
4.5mm Compression Plate (LC-LCP) (Narrow + Broad)	30	T-Locking Plate	12
Proximal Humerus Plate Anatomic	28	LCP Radius T-Plate, Bend angle	10
Proximal Femur Plate	28	LCP Superior Anterior Clavicle Plate	10
Proximal Tibia Lateral Plate	27	L-Buttress Plate	8
Distal Tibia Lateral Compression Plate	26	Distal Ulna Hook Plate	5
Calcaneus Plate	26		
Total (Plates)		689	
Nails	**Quantity**	**Cannulated screws**	**Quantity**
Cannulated Tibia Nail	32	Cannulated screw 4.5mm	50
Titanium Elastic Nail	30	Cannulated screw 3.5mm	42
Proximal Femoral Nail	29	Headlless Compression Screw	31
Long Reconstruction Femoral Nail	27	Cannulated screw 6.5mm	28
Humerus Nail	12	Cannulated screw 7.3mm	26
Total (Nails)	130	Total (Cannulared Screws)	177

Proximal humerus anatomic plate

The proximal humerus anatomic plate (Oltho medical) allows sending locking screws in various directions to improve stabilization and stabilize humeral head via rotator cuff sutures fixed on the plate through holes. 28 proximal humerus anatomic plates were used for the fixation of proximal humerus fractures. The union rate for the proximal humerus fractures was 100%. The implant survival rate for the proximal humerus plate until fracture union was 100%. No implant related complication was observed in study population.

Clavicle plates

Oltho medical clavicle plates include superior anterior locking compression clavicle plates and clavicle hook plates. Superior clavicle plate also has lateral extended version that allow sending locking 2.4mm screws to lateral fragments. A total of 34 clavicle plates were used in this study. The fracture union rates for clavicle shaft fractures was 100%. The implant survival rate for superior anterior locking compression clavicle plates till fracture union is 100%. No implant related complication was observed in the study population.

Oltho medical LCP clavicle hook plates allow treatment of distal clavicle fractures and acromioclavicular joint injuries. The healing rate for the distal clavicle fractures was 100%. The implant survival rate for the LCP clavicle hook plate till fracture union was 100%. No implant related complication was observed in study population.

10 patients received implant removal surgery at the 6 months follow up as routine elective implant removal.

Distal humerus plates

Oltho medical distal humerus plates included medial and lateral dorsal column plates used for fixation of distal humerus fractures. A total of 49 distal humerus medial and lateral dorsal plates were used for the fixation of distal humerus fractures. The union rate for the distal humerus fractures was 96%. There were no implant breakage in our medical records. Two fractures received revision fixation surgery with bone grafting, and both fractures healed uneventfully.

Olecranon plates

Oltho medical olecranon plates allow fixation olecranon fractures and olecranon osteotomies. 25 olecranon plates were analyzed. No implant breakage was recorded. The union rates for the olecranon fractures and olecranon ostetomies was 90%. Two patients received revision fixation surgery. Implant prominence and irritation was noted on 25% of the patients that is paralel to the literature.

Distal radius plates

Oltho medical distal radius plate included anatomic distal radius anatomic plates and LCP radius T-plate with bent angle. Anatomic distal radius plate are utilized for fixation of distal radius fractures. Locking distal screw configuration allows restoration of joint line and volar tilt. A total of 35 distal radius plates were analyzed. All fractures were healed uneventfully at the 3rd month follow up. No implant breakage was seen. Irritation of the extensor tendons was recorded due to the overextending screws in 3 patients (9%) which were removed following fracture union.

Proximal femur plates

Oltho medical proximal femur plates were used for the fixation of the proximal femur fractures, femoral osteotomies and trochanteric fractures. 28 proximal femur plates were analyzed. Union rate for the proximal femur fractures was 96%. One patient had trochanteric non-union which was asymptomatic and left untreated. No implant breakage was recorded.

Distal femur plates

Oltho medical distal femoral lateral plates were used for fixation of the distal femur fractures with or without cannulated screws. 26 distal femur plates were analyzed. The union rate for distal femur fractures were 100%. No implant breakage was observed.

Proximal tibia plates

Oltho medical proximal tibia plates included 4.5mm LCP medial plates and LCP lateral proximal tibia plates. A total of 48 proximal tibia plates were analyzed in 27 fractures. One patient suffered from non union that was successfully treated via revision fixation and bone grafting. Healing rate for the proximal tibia fractures was 96%. No implant breakage was noted.

Distal tibia plates

Oltho medical distal tibia plates included distal medial tibial plate and distal lateral compression plates. A total of 47 plates in 40 patients were analyzed. The union rate for distal tibia fractures were 92.5%. Three patients suffered from nonunion and received revision surgery. Screw breakage was seen in 2 (5%) patients due to nonunion, in 1 patient infection and soft tissue problems caused nonunion. No plate breakage was recorded.

Fibula plate

Thirty-two distal fibula anatomic plate was analyzed in 32 distal fibula fractures. The healing rate of the fibula fractures was 100%. No implant related problems were recorded. All of the implants survived throughout the treatment course.

Mini plates

Oltho medical Mini plates were used in various type of fracture fixation constructs like primary fixation providers in small bones as metatarsal and phalangeal fractures or supplementary fixation as anti rotation plates and neutralization plates. 36 mini plates, and a total of 37 L buttress, T buttress, and T locking plates were analyzed. There were no implant breakage recorded.

3.5 and 4.5 reconstruction plates

Oltho medical 3.5 and 4.5 reconstruction plates were used in primary or secondary fixation providers. A total of 53 reconstruction plates were analyzed and no implant failure was recorded in our study cohort.

3.5- 4.5 Compression plates [LC-DCP]

Oltho Medical 3.5 -4.5 compression plates provides compression of the fracture ends with cortical screws. Additionally adding locking screws to the construct would increase the rotational stability. A total of 65 LC-DCP plates were analyzed and no implant breakage was recorded.

Nails

A total of 130 intramedullary nails were analyzed in this study. Below are the specific data for the implants.

Femoral Nails: There was 1 nail breakage due to subtrochanteric fracture non-union and 1 interlocking screw breakage due to segmentary femur fracture out of 56 femoral nails that ended up with fracture union following implant removal and revision fixation surgery with the implant survival rates was 96.4% for Oltho medical femoral nails.

Tibia Nails: No nail breakage was recorded in 32 Oltho medical cannulated tibia nails. There were 2 interlocking screw breakages in 2 distal tibia fractures and 1 interlocking screw breakage tibia shaft breakage in an open tibia shaft fracture. Three cases needed exchange naill revision surgery all ended up with union.

Humerus nails: A total of 12 humerus intramedullary nails were analyzed and no implant breakage was observed in this study. All fractures healed uneventfully.

Titanium elastic nails: Oltho medical titanium elastic nails were utilized in pediatric trauma surgeries like tibia , femur and forearm fractures. Among 30 nails that were analyzed no implant breakage was recorded. All of the titanium elastic nails were removed following fracture union.

Cannulated screws and headless compression screws

Oltho medical cannulated screws and headless compression screws were used in the treatment of periarticular distal femur, proximal tibia, distal tibia, proximal femur, small bone fractures like scaphoid, metacarpals, phalangeal fractures foot fractures. Headless compression A total of 177 cannulated screws were analyzed and 1 implant breakage was recorded in a Lisfranc injury fixed with 4.0 headless cannulated screw. This implant breakage was attributed to early weight bearing.

Discussion

The present study demonstrated a high overall fracture union rate of 97.2% following surgical fixation, with implant failure observed in only 6 patients (1.3%). The overall implant survival rate of Oltho Medical trauma implants was 98.7%, indicating reliable biomechanical performance across a wide range of fracture types. Importantly, all fractures ultimately achieved union after appropriate management, underscoring the clinical effectiveness and safety of these implants in routine orthopedic trauma practice.

With every surgical procedure and implantation in orthopaedics there starts a race between implant failure and healing process of fracture [7]. Implant failure can be in form of plastic deformation, corrosion, or fatigue failure. Mechanical failure of implant can result due to repetitive loading and stress. In absence of union even the toughest metallic device and best designed implants are known to fail [13].

Technical aspects of implant failure have been studied in various studies [11,13] [22] [23]. According to laboratory results from two studies Azevedo in Brazil and Amel Farad H in Iran, manufacturers did not follow the standards in their cases of failed implant analysis. It is possible that implants differ in terms of purity and alloy inclusion from manufacturer to manufacturer [23]. In another study by Barbosa on three cases of implant failure, surgical technique and implant design, selection of implant have been reported to be important cause of implant failure [22]. Studies on implant quality control and designing errors are out of scope of this research and are limiting factors.

In study by Sharma et al. 2.4% implant failure was associated with deep infection[22]. In our study infection was found in 10 cases (2.2%). This infection was superficial to deep and associated with loosening of implant. Infection could have occurred during the surgery due to faulty surgical technique of open reduction and fixation of implant. Re-trauma is also major factor of implant failure during consolidation phase of fracture healing. Sharma et al. have also found re-trauma as significant cause for implant failure in their study.

Excessive body weight of the patient and early weight bearing on affected lower limb imparts more stress on implant during the healing stage of fracture. During the stance phase of gait cycle, load on lower limb is more than three times the body weight. Alfred O. Ogbemudia et al. in their study found patient non compliance and excessive body weight as significant reason for failure of implant and suggested cautious ambulation and graduated weight bearing [24]. Similar observations have been reported by Sharma et al., who emphasized that premature loading before sufficient callus formation increases the risk of screw loosening and fatigue failure [22]. Biomechanical analyses also demonstrate that repetitive cyclic loading under excessive weight significantly accelerates microcrack propagation within the implant material, particularly in stainless steel constructs [13,25]. In the present study, all instances of implant breakage occurred in patients with delayed union or early weight bearing, supporting these findings. Therefore, appropriate patient education, gradual rehabilitation, and close radiographic follow-up are essential to minimize mechanical overload and optimize implant longevity.

Plastic deformation, brittle and fatigue failure are known to occur with minor loads and re-trauma. 1.5% fatigue failure was observed in our study. Fatigue failure is associated with poor design, workman ship, handling and implant breaks from cyclical loading [7]. Surface notches or holes severely affect fatigue strength. Scratches or corrosion can also reduce the strength and predispose to implant failure. Fatigue failure of plates is more common than nails as intramedullary location of nails in shaft prevents some bending forces responsible for fatigue failure [13]. Plate ends act as stress riser leading to fresh fractures at ends. Plate fixation requires perfect reduction and anatomical reconstruction and may interfere with periosteal blood supply[26]. Poorly fixed implant with excessive soft tissue handling leads to failure of union and implant[26]. Recent biomechanical studies have demonstrated that even minor surface imperfections, such as micro-notches or corrosion pits, can initiate fatigue cracks and markedly reduce implant lifespan under repetitive cyclic loading [27]. Furthermore, comparative studies have shown that while plate constructs provide higher initial stiffness, intramedullary nails display superior fatigue endurance due to their load-sharing nature and reduced exposure to bending stresses [28]. These findings are in accordance with our observations, suggesting that implant longevity depends not only on material quality but also on mechanical design and surgical technique.

Intra-medullary implants are load sharing and provide good stability to fractures of long bone, allowing early rehabilitation and functional recovery of patient. Locked intra-medullary nail provide excellent axial and rotational stability as compared to unlocked intra-medullary nail. Intra-medullary implant failures occur with small diameter nail, improper selection of implant which is not suitable for unstable fracture configurations. Failure to provide rotational stability at fracture site leads to loosening of implant and failure [11]. At times inability to select a suitable implant to match the fracture configuration or an improper surgical technique to restore fracture anatomy, inadequate fixations, pre/post-operative complications like infections, patient non-compliance with implant instructions and degree of union lead to failure of implant [25].

We identified 1 cephalomedullary nail breakage in a subtrochanteric femur fracture 12 months after the surgery caused by nonunion of the fracture. The nail breakage occurred in the shaft-neck intersection and this type of failure is typically seen because of metal fatigue due to excessive loading following fracture non-union. In 2 patients with distal tibia fracture, broken screws of the distal medial tibia plate was observed. This was attributed to early weight bearing on the extremity and extended fracture union time. Broken interlocking screws of the femoral nail was noted 1 one, tibial nail in 1 patient. Interlocking screws of the femoral and tibia nails occured following delayed union and cause auto-dynamization of the fracture. Both fractures healed uneventfully. Additionally, 1 4mm headless compression screw was broken following surgical treatment of Lisfranc injury at the 6 months follow-up. The screw breakage was attributed to the motion between fixed fragments due to weight bearing. No implant removal was required.

The major limitation of this study was its retrospective nature. Prospective randomized studies would be the ideal way of determining implant success in trauma surgery. However, fracture healing rates with a minimum of 6 months follow-up would provide sufficient information for survival analysis. Additionally, this study lacks clinical outcomes. In trauma surgery, clinical outcomes are more related to the fracture type, injury pattern, patient specific variables like age, associated comorbidities and activity levels. Thus, analyzing clinical outcomes were beyond the scope of this study.

Conclusion

In conclusion medical trauma implants provided good to excellent survivability compared to the literature till fracture healing. Even though complications were noticed during follow-ups all were successfully managed with appropriate approach.

Author contribution

Conception: S.K.; Design: S.K.; Data acquisition: M.O., U.C.K.; Data analysis: M.O., U.C.K.; Data interpretation: U.C.K.; Drafting of the manuscript: M.O.; Critical revision of the manuscript: S.K. All authors reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of this study.

Ethical approval

This study was approved by the Hacettepe University Health Sciences Research Ethics Committee (Date: 02.05.2022, Decision/Protocol No: IRB: GO22/942). Informed consent was obtained from all participants involved in this study.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The authors declare that this study received no funding.

Generative AI statement

The authors declare that no generative AI or AI-assisted technologies were used in the writing or preparation of this study.

References

[1] Einhorn TA, Gerstenfeld LC. Fracture healing: mechanisms and interventions. Nat Rev Rheumatol 2015;11(1):45-54. **[Crossref]**

[2] Phillips AM. Overview of the fracture healing cascade. Injury 2005;36(Suppl 3):S5-7. **[Crossref]**

[3] Marsell R, Einhorn TA. The biology of fracture healing. Injury 2011;42(6):551-5. **[Crossref]**

[4] Bishop JA, Palanca AA, Bellino MJ, Lowenberg DW. Assessment of compromised fracture healing. J Am Acad Orthop Surg 2012;20(5):273-82. **[Crossref]**

[5] Giannoudis PV, Jones E, Einhorn TA. Fracture healing and bone repair. Injury 2011;42(6):549-50. **[Crossref]**

[6] Rath B, Nam J, Knobloch TJ, Lannutti JJ, Agarwal S. Compressive forces induce osteogenic gene expression in calvarial osteoblasts. J Biomech 2008;41(5):1095-103. **[Crossref]**

[7] Carter DR, Beaupré GS, Giori NJ, Helms JA. Mechanobiology of skeletal regeneration. Clin Orthop Relat Res 1998;355:S41-55. **[Crossref]**

[8] Shapiro F. Cortical bone repair. The relationship of the lacunar-canalicular system and intercellular gap junctions to the repair process. J Bone Joint Surg Am 1988;70(7):1067-81.

[9] Gerstenfeld LC, Alkhiary YM, Krall EA, et al. Three-dimensional reconstruction of fracture callus morphogenesis. J Histochem Cytochem 2006;54(11):1215-28. **[Crossref]**

[10] Perren SM. Evolution of the internal fixation of long bone fractures. The scientific basis of biological internal fixation: choosing a new balance between stability and biology. J Bone Joint Surg Br 2002;84(8):1093-110. **[Crossref]**

[11] Kuzyk PR, Schemitsch EH. The basic science of peri-implant bone healing. Indian J Orthop 2011;45(2):108-15. **[Crossref]**

[12] Shaikh M, Kahwash F, Lu Z, Alkhreisat M, Mohammad A, Shyha I. Revolutionising orthopaedic implants-a comprehensive review on metal 3D printing with materials, design strategies, manufacturing technologies, and post-process machining advancements. The International Journal of Advanced Manufacturing Technology 2024;134(3):1043-76. **[Crossref]**

[13] Kim T, See CW, Li X, Zhu D. Orthopedic implants and devices for bone fractures and defects: past, present and perspective. Engineered Regeneration 2020;1:6-18. **[Crossref]**

[14] Li Z, Kuhn G, von Salis-Soglio M, et al. In vivo monitoring of bone architecture and remodeling after implant insertion: the different responses of cortical and trabecular bone. Bone 2015;81:468-477. **[Crossref]**

[15] Zura R, Xiong Z, Einhorn T, et al. Epidemiology of fracture nonunion in 18 human bones. JAMA Surg 2016;151(11):e162775. **[Crossref]**

[16] Hammer RR, Hammerby S, Lindholm B. Accuracy of radiologic assessment of tibial shaft fracture union in humans. Clin Orthop Relat Res 1985;(199):233-8.

[17] Nicholson JA, Yapp LZ, Keating JF, Simpson AHRW. Monitoring of fracture healing. Update on current and future imaging modalities to predict union. Injury 2021;52(Suppl 2):S29-34. **[Crossref]**

[18] Schwarzenberg P, Darwiche S, Yoon RS, Dailey HL. Imaging modalities to assess fracture healing. Curr Osteoporos Rep 2020;18(3):169-179. **[Crossref]**

[19] Whelan DB, Bhandari M, McKee MD, et al. Interobserver and intraobserver variation in the assessment of the healing of tibial fractures after intramedullary fixation. J Bone Joint Surg Br 2002;84(1):15-8. **[Crossref]**

[20] Leow JM, Clement ND, Tawonsawatruk T, Simpson CJ, Simpson AHRW. The radiographic union scale in tibial (RUST) fractures: reliability of the outcome measure at an independent centre. Bone Joint Res 2016;5(4):116-21. **[Crossref]**

[21] Whelan DB, Bhandari M, Stephen D, et al. Development of the radiographic union score for tibial fractures for the assessment of tibial fracture healing after intramedullary fixation. J Trauma 2010;68(3):629-32. **[Crossref]**

[22] Sharma AK, Kumar A, Joshi GR, John JT. Retrospective study of implant failure in orthopaedic surgery. Med J Armed Forces India 2006;62(1):70-2. **[Crossref]**

[23] Peivandi MT, Yusof-Sani MR, Amel-Farzad H. Exploring the reasons for orthopedic implant failure in traumatic fractures of the lower limb. Arch Iran Med 2013;16(8):478-82.

[24] Ogbemudia AO, Umebese PFA. Implant failure in osteosynthesis of fractures of long bones. Journal of Medicine and Biomedical Research 2009;5(2). **[Crossref]**

[25] Chen CE, Weng LH, Ko JY, Wang CJ. Management of nonunion associated with broken intramedullary nail of the femur. Orthopedics 2008;31(1):78. **[Crossref]**

[26] Beeharry MW, Ahmad B. Principles of fracture healing and fixation: a literature review. Cureus 2024;16(12):e76250. **[Crossref]**

[27] Wu K, Li B, Guo JJ. Fatigue crack growth and fracture of internal fixation materials in in vivo environments-a review. Materials (Basel) 2021;14(1):176. **[Crossref]**

[28] Kandemir U, Herfat S, Herzog M, Viscogliosi P, Pekmezci M. Fatigue failure in extra-articular proximal tibia fractures: locking intramedullary nail versus double locking plates-a biomechanical study. J Orthop Trauma 2017;31(2):e49-54. **[Crossref]**

acta medica

RESEARCH ARTICLE

Impact of age and solvent exposure on audiometric abnormalities among automotive production workers: A cross-sectional study from Türkiye

Fatma Bozdağ[1], Sultan Pınar Çetintepe[1], Volkan Medeni[1], Mustafa Necmi İlhan[1]

[1]Division of Occupational Medicine, Department of Public Health, Faculty of Medicine, Gazi University, Ankara, Türkiye

Abstract

Objectives: This study investigates the relationship between occupational noise, solvent exposure, and audiometric abnormalities in male automotive industry workers, focusing on factors associated with audiometric abnormalities, including age, body mass index, and biochemical markers.

Methods: This cross-sectional study was conducted among 500 male workers who were admitted to the outpatient clinic of Ankara Occupational and Environmental Diseases Hospital for periodic health examinations between September 1, 2020, and July 1, 2021. Audiometric results and biochemical markers, including hippuric acid, trichloroacetic acid, and manganese, were assessed alongside demographic and occupational variables. Multivariate logistic regression was used to identify predictive factors for abnormal audiometry results.

Results: Among the participants, 61.2 % exhibited abnormal audiometry results, and 11.4 % had noise-induced hearing loss. Age strongly predicted abnormal audiometry (OR = 1.11, 95% CI: 1.07–1.15, $p < 0.001$), with an 11% increase in odds per year. Elevated trichloroacetic acid (TCA) levels were also associated with a 9% increase in odds of abnormal audiometry (OR = 1.09, 95% CI: 1.01–1.18, $p = 0.027$). No significant associations were found for body mass index, duration of work, or hippuric acid levels.

Conclusions: Age and TCA levels were significantly associated with audiometric abnormalities among workers exposed to occupational noise and solvents. These findings highlight the need for targeted interventions, including regular audiometric evaluations and minimizing solvent exposure, to reduce the risk of hearing loss in industrial settings. Further research is needed to explore the underlying mechanisms linking solvent exposure and auditory dysfunction.

Keywords: noise-induced hearing loss, occupational solvent exposure, audiometric abnormalities, automotive industry workers, Trichloroacetic Acid (TCA), predictive risk factors

Introduction

Occupational noise is a common physical hazard in many industrial workplaces. It is mainly generated by machinery, metal processing, and other manufacturing activities, and workers may be exposed to sound levels that exceed recommended safety limits during routine operations. To reduce the health risks associated with excessive noise exposure, several international organizations have established regulatory frameworks and exposure limits for occupational settings. For instance, the Occupational Safety and Health Administration (OSHA) and the National Institute for Occupational Safety and Health (NIOSH) recommend

Corresponding author: Fatma Bozdağ • **Email:** fatmabozdag7@gmail.com

Received: October 10, 2025 **Accepted:** April 13, 2026 **Published online:** June 28, 2026

exposure limits of approximately 85–90 dB over an 8-hour work shift, with preventive measures required when these limits are exceeded [1,2]. Prolonged exposure to high levels of occupational noise may lead to noise-induced hearing loss (NIHL), a progressive sensorineural hearing disorder that typically begins at high frequencies. Studies show the relationship between workplace exposure to noise and noise-induced hearing loss (NIHL). Over time, prolonged exposure to high sound levels can cause permanent inner ear damage, called noise-induced hearing loss. This sensorineural hearing loss begins in the high-frequency range and progresses gradually with continued exposure to extreme sound levels [3]. It has been found in animal experiments that heavy metals and organic solvents are ototoxic [4-6]. Various chemicals with ototoxic properties have been identified, and organic solvents have been shown to increase hearing loss synergistically with noise [7]. In a study conducted among printing workers, it was found that there was a significant relationship between hippuric acid level and hearing loss [8]. The concentrations of methyl hippuric acid in urine and pure-tone thresholds (2 to 8 kHz) were positively correlated in xylene-exposed workers. In the same study, subjects with a high cumulative dose of xylene exposure had poorer test results than subjects with a low cumulative xylene exposure [9].

In the literature, there are studies in the form of case reports supporting that chronic exposure to manganese may be associated with hearing impairment [10,11]. A mouse study determined that the level of manganese accumulated in the inner ear of the subject who was injected subcutaneously for three days and remained in the inner ear for at least 2 weeks after the treatment. This is important as a study demonstrates manganese accumulation in the inner ear due to systemic exposure [12].

After diagnosing occupational diseases in Turkey, it is mandatory to report them to SGK (Sosyal Güvenlik Kurumu, Social Security Institution). According to the latest published 2021 SGK Occupational Diseases and Occupational Accidents Data, the total number of reported occupational diseases in Turkey is 1207, 55 of which have been reported as occupational hearing loss [13]. A significant relationship was found between age, duration of exposure to noise at work, and hearing loss in forestry workers using chainsaws in Turkey [14]. In another case-control study, no correlation was found between blood arsenic levels and hearing loss levels in miners [15]. It has been found that there is a loss of capacity for the vestibular-ocular reflex, which shows vestibular dysfunction in patients with hearing loss due to exposure to noise in industrial workers [16]. Work-related noise and hearing loss were also associated with carpenters [17].

According to the data for 2021, there are 16 million insured employees in Turkey, 238,000 of whom work in the automotive production sector [13]. In Turkey, workers in automobile manufacturing plants typically work 5 days a week, at least 8 hours a day. The total weekly working time is at most 45 hours. The automobile manufacturing sector has main occupational groups such as painters, assemblers, and welders [18]. In the production sector, the sound of metals hitting and rubbing against each other creates noise in the work environment. Paint, solvents, and welding fumes, used extensively in the automobile industry, can also cause inner ear damage, as mentioned above [4,11].

Previous research has found that these chemicals can cause damage to the inner ear. However, little evidence suggests a relationship between workplace manganese and solvent exposure and hearing. Studies on the effect of paint and welding fume exposure on hearing in the automotive manufacturing sector are minimal. Therefore, this study aims to determine the relationship between urinary manganese levels, solvent degradation metabolites (urine TCA, hippuric acid), and audiometric abnormality levels among automotive manufacturing workers.

Methodology

Study design

This cross-sectional study examines the patient files of the automotive factory employees who applied to the occupational outpatient clinic of Gazi Hospital for periodic health examinations between September 1, 2020, and July 1, 2021. xx University Human Research Ethics Committee granted the study exemption from full ethical approval (approval no: 20.10.2022-E-77082166). All the workers included in the study are men in different occupations in the same automotive manufacturing factory. The entire universe is included in the study. An otolaryngologist examined all employees, and those with chronic otitis media were excluded from the study. Individuals with personal or familial deafness

in their anamnesis, those who had ear surgery, those who used ototoxic drugs, and those with type 2 diabetes mellitus and hypertension for more than 5 years were excluded from the study. Workplace noise levels were not measured. Workplace noise levels were not directly measured in this study. However, all participants were employed in the automotive production environment, where routine processes such as metal processing, welding, and assembly are known to generate substantial occupational noise. Therefore, workers were considered to be exposed to workplace noise based on their job roles and the characteristics of the production environment. According to the audiometry results, the groups were divided into those with normal and abnormal audiometry results. To diagnose NIHL, Klockhoff-modified criteria for a history of occupational noise exposure, bilateral hearing loss, and a threshold level greater than 25 dB at frequencies between 1,000 Hz and 8,000 Hz were used [19].

Data collection

Age, gender, height, weight, BMI (Body Mass Index), current and past occupations, working years, workplace exposures, smoking and alcohol use status, chronic diseases and drug use, previous ear diseases, and operations were obtained from the medical anamnesis form. The employees' audiometer analysis reports, hemogram, biochemistry, and toxicological marker results were examined from the patient files. Due to the small number of participants in some occupational categories (assembler and quality control officer), these groups were not included as separate categories in the comparative analyses. Body mass index (BMI) was calculated by dividing body weight in kilograms by the square of height in meters (kg/m^2), according to the standard anthropometric definition recommended by the World Health Organization [20]. The employees' audiometry reports, hemogram results, biochemical parameters (glucose and thyroid-stimulating hormone [TSH]), and toxicological markers (urinary hippuric acid, TCA], and whole-blood manganese levels) were obtained from patient records and included in the analysis.

Audiometric examination

An otolaryngologist examined all of the employees after a pure tone audiometry was performed using a pure tone manual diagnostic audiometer (Model GSI 61, Grason-Stadler, Inc.) by a single audiologist at the Audiology Laboratory of Ankara Occupational and Environmental Diseases Hospital. Pure tone audiometry was conducted with the subjects at frequencies of 0.5, 1, 2, 3, 4, and 6 kHz using air and bone conduction in a sound-isolated chamber. The participants were informed to differentiate between low sound levels of several frequency pure tones and react by pressing a button. The lowest tone heard at each frequency was evaluated as the hearing threshold level. A normal audiogram was described as findings consistent with normal hearing sensitivity, typically consisting of air conduction and bone conduction shown for the right and left ear, respectively, with <25 dB HL values at each tested frequency level [21]. In this study, "abnormal audiometry" referred to any audiometric finding deviating from normal hearing thresholds. In contrast, NIHL was defined using the Klockhoff-modified criteria, which identify hearing loss patterns compatible with occupational noise exposure. Therefore, NIHL represents a specific subset of abnormal audiometric findings. Workplace environmental noise measurements were not performed.

Collection of biological samples

Ten millilitres of venous blood were poured into tubes (BD Vacutainer, USA) by venipuncture from each subject and processed within the following 3 days. After centrifuging the samples at 3,500 g for 10 minutes at +4 °C, the serum was separated and stored at -80 °C until analysis. Whole blood was used for manganese analysis. Urine was collected into a clean urine container. After collecting the urine, it was stored in the refrigerator at +4°C. It was sent to the analytical toxicology laboratory for analysis and frozen before analysis.

Biochemical measurements

Complete blood counts were analysed using the Coulter Gen-S haematology analyser. Whole blood manganese was analysed with certified reference material (Seronorm Trace Elements Billinstod, Norway) by Inductively Coupled Plasma—Mass Spectrometer on Agilent 7700 (Agilent Technologies, USA). Urine hippuric acid rates were analysed using "Chromosystems, Agilent 1200 series equipped with a UV detector, HPLC with reagent kit for the HPLC-Analysis. Urinary TCA was measured with a commercial kit (FAR, Verona, Italy). Briefly, TCA in the sample reacts with pyridine in an alkali medium and forms a colour complex that can be photometrically determined at 526 nm. The assay principle is the colorimetric endpoint.

Statistical analysis

Statistical analyses were performed using the SPSS software version 26 (Statistical Package for the Social Sciences by IBM). The variable was investigated using visual (histograms, probability plots) and analytical methods (Kolmogorov-Smirnov/Shapiro-Wilk's test) to determine whether or not they are typically distributed. Descriptive analyses were presented using medians for the non-normally distributed and ordinal variables. Since the age, weight, height, BMI, duration of work, hippuric acid, manganese, TCA, TSH, and glucose levels were not normally distributed, nonparametric tests were conducted to compare these parameters and the ordinal variables. The Mann-Whitney U test was used to compare the differences between workers with normal and abnormal audiometry for age, weight, height, BMI, duration of work, hippuric acid, manganese, TCA, TSH, and glucose levels. The proportions of workers with abnormal and standard audiometry were presented by smoking status and occupations using cross-tabulations. The Chi-square test was used to compare these proportions in different groups. Multivariate logistic regression was employed to assess the independent effects of predictors (age, BMI, duration of work, hippuric acid, and TCA levels) on the likelihood of abnormal audiometry results, with odds ratios (OR) and 95% confidence intervals (CI) calculated to quantify associations, and p-values used to determine statistical significance. Multicollinearity among the independent variables included in the regression model was assessed using the variance inflation factor (VIF). All VIF values were below the commonly accepted threshold, indicating no significant multicollinearity.

Results

A total of 500 male workers employed in the automotive production sector were included in the study. The median age of the participants was 33 years, and the median body mass index (BMI) was 26,1 kg/m^2. More than half of the workers (62.4%) were current smokers. The median duration of employment in the sector was 6 years. The most common occupations were welding (50.8%) and painting (43.8%). Overall, abnormal audiometry findings were observed in 61.2% of the workers. The detailed sociodemographic and biochemical characteristics of the study population are presented in Table 1. All subjects in the study group recruited in the automotive industry were male. The median age of the subjects was 33. The median height was 175 cm, and the median weight was 80 kilograms. The median BMI calculated with the kg/m^2 formula was determined as 26.1. More than half of the employees were smokers (62.4%). While the maximum working period of the employees in the same sector is 17 years, the average working year has been determined to be 6 years. While reviewing the occupations of the employees, it was determined that 56.2% were welders and 43.8% were painters. When the data of all employees were examined, it was determined that 61.2% of the individuals had abnormal findings in audiometry. In toxicological analyses, the median value of hippuric acid was 521.5 mg/L, the median value of TCA was 4.4 µg/L, and the median value of manganese was 9.9 µg/L among all workers. In biochemical analyses, the median glucose level was 96 mg/dl, and the median TSH level was 1.86 mIU/L.

When the differences between the group with abnormal audiometry and the group with normal audiometry results were examined, age was statistically significantly

Table 1. Sociodemographic and biochemical characteristics of male industry workers

Age (years), median (Q1–Q3)		33 (27-43)
Height (cm), median (Q1–Q3)		175 (171-180)
Weight (kg), median (Q1–Q3)		80 (74-90)
BMI (kg/m^2), median (Q1–Q3)		26.1 (24.2-28.7)
Current smoking, n (%)		312 (62.4)
Duration of work (years), median (Q1–Q3)		6.0 (4-17)
Occupation, n (%)	Welder	254 (50.8)
	Painter	219 (43.8)
	Assembler	22 (4,4)
	Quality control officer	5 (1,0)
Abnormal audiometry, n (%)		306 (61.2)
NIHL, n (%)		57 (11.4)
Hippuric acid (mg/L), median (Q1–Q3)		528 (295-924)
TCA (µg/L), median (Q1–Q3)		4.4 (3.2-6.3)
Manganese (µg/L), median (Q1–Q3)		9.8 (7.9-12.4)
Glucose (mg/dL), median (Q1–Q3)		96 (90-104)
TSH (mIU/L), median (Q1–Q3)		1.79 (1.29-2.52)

N: number, BMI: body mass index, NIHL: Noise-Induced Hearing Loss, TCA: Trichloroacetic Acid, TSH: Thyroid-Stimulating Hormone, mg/L: Milligrams per Liter, µg/L: Micrograms per Liter, mIU/L: Milli-International Units per Liter.

higher in the group with abnormal audiometry results (p<0.001). The BMI value in the group with abnormal audiometry results is higher than in the group with standard audiometry. This difference was also found to be statistically significant (p=0.009). The groups had no significant difference regarding height, weight, and smoking status. There is a statistically significant difference between the groups regarding working years in favor of those with abnormal audiometry (p<0.001). There was no significant difference between the groups in terms of occupations. Hippuric acid and TCA levels were statistically significantly higher in the group with abnormal audiometry (p=0.015, p=0.040, respectively). There was no significant difference between the groups regarding manganese, glucose, and TSH levels (Table 2).

When the differences between those with and without sensorineural hearing loss were examined, the median age was 40 in those with sensorineural hearing loss and higher than in those with no NIHL. This difference is statistically significant. Those with NIHL had a statistically higher BMI than those without. There was no significant difference between the groups regarding height and weight. When evaluated in terms of smoking, 54.2% of those with NIHL and 64.4% of the No NIHL audiometry group were smokers. This difference is not statistically significant. Considering the median working years, this period was 8 years in patients with sensorineural hearing loss, statistically significantly higher than in the group with other audiometry. The groups showed no significant difference in the hippuric acid, TCA and manganese median. The median TSH level was 1.76 mIU/L and statistically significantly lower in the NIHL group. There was no significant difference between the groups regarding glucose levels (Table 3).

The factors associated with abnormal audiometry results among male industry workers exposed to occupational noise are presented in Table 4. Multivariate logistic regression analysis showed that age and TCA levels were significantly associated with abnormal audiometry results. Age was strongly associated with abnormal audiometry results, with an odds ratio (OR) of 1.11 (95% CI: 1.07–1.15, p < 0.001), indicating that for every one-year increase in age, the odds of having abnormal audiometry results increased by 11%. Similarly, TCA levels were significantly associated with abnormal audiometry results, with an OR of 1.09 (95% CI: 1.01–1.18, p = 0.027), suggesting that for each unit increase in TCA concentration, the odds of abnormal audiometry results increased by 9 %. Other variables, including body mass index (BMI) (OR = 1.01, 95% CI: 0.96–1.07, p = 0.626), duration of work (OR = 0.98, 95% CI: 0.94–1.02, p = 0.448), and hippuric acid levels (OR = 1.00,

Table 2. Comparison of characteristics according to audiometry results

		Audiometry		
Variable		**Normal**	**Abnormal**	**p-value**
Age (years), median (Q1–Q3), (n=500)		28.5 (26.0-36.3)	39.0 (29.0-45.0)	**<0.001***
BMI (kg/m^2), median (Q1–Q3), (n=496)		25.7 (24.0-28.0)	26.8 (24.3-29.0)	**0.018***
Smoking status	Current smoker, n (%), (n=312)	128 (41.0)	184 (59.0)	0.188†
	Nonsmoker, n (%), (n=188)	66 (35,1)	122 (64,9)	
Occupation	Welder, n (%), (n=254)	99 (51.0)	155 (50.7)	0.810†
	Painter, n (%), (n=219)	83 (42.8)	136 (44.4)	
Duration of work (years), median (Q1–Q3), (n=500)		5 (3-8)	7 (4-20)	**<0.001***
Hippuric acid (mg/L), median (Q1–Q3), (n=485)		453.0 (282.0-851.0)	600.0 (322.0-991.5)	**0.015***
TCA (µg/L), median (Q1–Q3), (n=489)		4.2 (3.1-6.2)	4.7 (3.4-6.5)	**0.040***
Manganese (µg/L), median (Q1–Q3), (n=486)		9.7 (8.0 -12.6)	9.8 (7,9-12.3)	0.757*
Glucose (mg/dL), median (Q1–Q3), (n=490)		96 (90-101)	96 (90-105)	0.250*
TSH (mIU/L), median (Q1–Q3), (n=470)		1.75 (1.31-2.50)	1.80 (1.27-2.54)	0.866*

For significance testing, a Chi-square† was formed for the differences in group numbers, and a Mann-Whitney-U* was performed.
N: number, BMI: body mass index, NIHL: Noise-Induced Hearing Loss, TCA: Trichloroacetic Acid, TSH: Thyroid-Stimulating Hormone, mg/L: Milligrams per Liter, µg/L: Micrograms per Liter, mIU/L: Milli-International Units per Liter.

95% CI: 1.00–1.00, p = 0.365), did not show statistically significant associations with abnormal audiometry results in this model.

Discussion

It was observed that audiometry evaluations were not within normal limits in 61.2% of the workers in our study. Hearing loss is present in 19.2% of all participants. In a study evaluating the hearing level of workers exposed to noise and organophosphates in the workplace, an abnormality was found in the audiometry results of 53.1% of the participants [22]. In a study conducted among employees of 22 automobile factories, the prevalence of noise-induced hearing loss was 28.8% [23]. A survey conducted to determine the level of hearing health and noise exposure among workers in the automotive industry determined that 26.8% of the participants had various levels of hearing loss [18]. In a study conducted on 2647 people working in an automobile factory, the prevalence of high-frequency hearing loss was 17.2% among the participants [24]. All these results show that a significant portion of the

Table 3. Comparison of characteristics according to the presence of NIHL

		Audiometry		
Variable		**No NIHL**	**NIHL**	**p-value**
Age (years), median (Q1–Q3), (n=500)		32 (27-43)	40 (29-45)	0.016*
BMI (kg/m^2), median (Q1–Q3), (n=496)		26.0 (24.2-28.4)	27.5 (24.7-29.2)	0.083*
Smoking status	Current smoker, n (%), (n=312)	281 (90.1)	31 (9.9)	0.184†
	Non-smoker, n (%), (n=188)	162 (86,2)	26 (13,8)	
Occupation	Welder, n (%), (n=254)	230 (90.6)	24 (9.4)	0.147†
	Painter, n (%), (n=219)	189 (86.3)	30 (13.7)	
Duration of work (years), median (Q1–Q3), (n=500)		5 (3-16)	8 (6-20)	0.006*
Hippuric acid (mg /L), median (Q1–Q3), (n=485)		525.0 (294.0-925.0)	585.0 (319.8-909.8)	0.891*
TCA (µg/L), median (Q1–Q3), (n=489)		4.5 (3.3-6.4)	4.3 (2.8-6.0)	0.145*
Manganese (µg/L), median (Q1–Q3), (n=486)		9.8 (7.9-12.5)	9.7 (7.9-11.7)	0.480*
Glucose (mg/dL), median (Q1–Q3), (n=490)		96 (90-104)	97 (89-103)	0.927*
TSH (mIU/L), median (Q1–Q3), (n=470)		1.80 (1.31-2.52)	1.73 (1.11-2.65)	0.438*

For significance testing, a Chi-square† was formed for the differences in group numbers, and a Mann-Whitney-U* was. N: number, BMI: body mass index, NIHL: Noise-Induced Hearing Loss, TCA: Trichloroacetic Acid, TSH: Thyroid-Stimulating Hormone, mg/L: Milligrams per Liter, µg/L: Micrograms per Liter, mIU/L: Milli-International Units per Liter.

Table 4. Predictive factors of abnormal audiometry results among male industry workers exposed to occupational noise

Variable	**OR (% 95% CI)**	**p-value**
Age	**1.11 (1.07-1.15)**	**<0.001**
BMI (kg/m^2)	1.01 (0.96-1.07)	0.626
Duration of work (years)	0.98 (0.94-1.02)	0.448
Hippuric acid (mg/L)	1.00 (1.00-1.00)	0.365
TCA (µg/L)	**1.09 (1.01-1.18)**	**0.027**

Multivariate logistic regression was used to calculate the means for each category. BMI: body mass index, TCA Trichloroacetic Acid, mg/L Milligrams per Liter, µg/L Micrograms per Liter.

workers in the automotive industry have hearing loss. Our findings are compatible with the literature. The differences between the frequencies may be due to the lack of similarity in characteristics such as age, gender, working unit, noise exposure, and personal protective equipment use of the individuals in the sample.

Our results showed that age and working time were statistically significantly higher in participants with abnormal audiometry results and those with noise-induced hearing loss. In our study, the odds of abnormal audiometry results increased by 11% for every one-year increase in age. In a survey of workers in automobile parts manufacturing factories, the frequency of hearing loss development in any ear is 33.3% for the participants with a working period of 10 years or more and 20.6% for those with less than 10 years. This frequency, which was 12.2% in the 20-30 age group, was determined as 27.5% in the 31-50 age group [25]. In a study investigating the effectiveness of earplugs in preventing noise-induced hearing loss in an automobile factory, it was observed that the hearing thresholds of workers exposed to noise increased significantly as age and exposure time increased [26]. A study investigating the characteristics of hearing loss in workers exposed to noise in the automotive manufacturing industry revealed that hearing loss is associated with working time [27]. In a prospective study that determined the 10-year incidence of workplace noise exposure and age-related hearing loss, evidence was found that age and exposure time have a potential multiplier effect on hearing function [28]. Undoubtedly, the findings in our study can be considered significant, considering that hearing loss occurs due to chronic exposure to noise and cumulative cochlear microtraumas [29]. In this study, increasing age was associated with abnormal audiometric findings. However, hearing impairment may also occur as part of the natural aging process (presbycusis), even in individuals without occupational exposures. Although NIHL was defined using the Klockhoff-modified criteria to identify audiometric patterns compatible with occupational noise exposure, the contribution of age-related hearing changes cannot be completely ruled out. Therefore, age should be considered a potential confounding factor when interpreting the association between occupational exposures and hearing impairment.

In this study, the mean body mass index of workers with abnormal audiometry results and noise-induced hearing loss was higher than that of workers with standard audiometry. In a study that included 48,549 employees aged 20-64 who did not have hearing loss at baseline and aimed to prospectively determine the link between obesity and hearing loss, overweight and obesity were found to be associated with an increased risk of hearing loss [30]. The findings of a meta-analysis to clarify further the potential relationship between body mass index and hearing loss contributed to the evidence that high body mass index may be positively associated with the risk of hearing loss [31]. In a longitudinal study of 249,000 Swedes that determined the estimated relative risks for sensorineural hearing loss, obesity was associated with doubling the risk of hearing loss [31]. Many different mechanisms can explain the relationship between obesity and hearing loss. Obesity can cause hearing loss by directly damaging hair cells through oxidative stress [32]. Obesity-related atherosclerosis may reduce blood flow to the cochlea by causing the narrowing of the artery and may result in decreased auditory sensitivity [8,33]. In addition, obesity-induced inflammation may contribute to target organ damage, and degeneration in hearing may be exacerbated by increased hypoxia and loss of spiral ganglion cells [28,34].

In this study, workers with hearing loss showed lower TSH levels compared with those with normal audiometry results. However, TSH levels alone are insufficient to determine the presence or type of thyroid dysfunction, as decreased TSH levels may also reflect hyperthyroidism. Since free T3 and free T4 measurements were not available in this study, the relationship between thyroid function and hearing impairment should be interpreted cautiously. Further studies including comprehensive thyroid hormone assessments are needed to better clarify this potential association.

In our study, hippuric acid and TCA levels were higher in participants with abnormal audiometry. TCA levels were significantly associated with abnormal audiometry results, suggesting that for each unit increase in TCA concentration, the odds of abnormal audiometry results increased by 9 %. Hippuric acid and TCA are the primary metabolites of toluene, and trichloroethylene is among the most essential organic solvents [33]. A study evaluating hearing impairments in workers exposed to various levels of noise and solvents in a printing facility showed that occupational exposure to toluene caused an increase in hippuric acid, a nonspecific metabolite, in the urine. A rise of one gram of hippuric acid in the urine increased the probability of hearing loss by 1.76 times

[8]. Results of a study of workers exposed to solvents at concentrations below the occupational exposure limit values indicate that exposure to even moderate concentrations of occupational organic solvents increases the risk of hearing loss [35]. In a five-year study documenting the potential development of toluene and noise-induced hearing loss through repeated testing and measurement, the estimated relative risk of elevated hippuric acid levels for high-frequency hearing loss was 1.28 [34]. In a relatively old study of workers with chronic trichloroethylene poisoning, hearing loss was observed in 26 of 40 workers who were frequently exposed to high concentrations of trichloroethylene. Hearing loss in workers with urinary TCA levels between 40-200 mg/l was bilateral, affecting sensorineural and high frequencies [36]. In an animal study in four groups of rats exposed to trichloroethylene alone, the noise alone, a combination of trichloroethylene and noise, or control conditions, hearing loss due to exposure to a combination of trichloroethylene and noise at a frequency of 4 kHz was significantly greater than in the other groups [37]. The need for current data on the effects of solvent exposure on hearing function is noteworthy in the literature.

Limitation

This study has several limitations that should be acknowledged when interpreting the results. First, workplace noise levels were not directly measured using environmental noise monitoring; therefore, individual noise exposure levels could not be quantified precisely. Second, information on the use of personal hearing protection devices was not available, which may have influenced the audiometric outcomes observed among the workers. Third, due to the cross-sectional design of the study, causal relationships between solvent exposure, occupational noise, and hearing impairment cannot be established. Finally, although several relevant variables were included in the analysis, the possibility of residual confounding from unmeasured factors cannot be entirely ruled out.

Conclusion

This study highlights age and TCA levels significantly associated with audiometric abnormalities among male automotive industry workers exposed to occupational noise and solvents. The findings demonstrate an 11% increase in the odds of abnormal audiometry per year of age and a 9% increase per unit rise in TCA levels, emphasizing the cumulative impact of aging and chemical exposure on auditory health. In contrast, body mass index (BMI), duration of work, and hippuric acid levels did not show significant associations. These results underscore the importance of integrating regular audiometric evaluations, targeted interventions to reduce solvent exposure, and hearing conservation programs to mitigate the risks of occupational hearing loss in industrial settings. Further research is needed to explore the mechanisms linking solvent exposure to auditory dysfunction and develop evidence-based preventive strategies.

Author contribution

Conception: F.B., S.P.Ç., V.M., M.N.İ.; Design: F.B., S.P.Ç., M.N.İ.; Data acquisition: S.P.Ç.; Data analysis: F.B., S.P.Ç., V.M.; Data interpretation: S.P.Ç.; Drafting of the manuscript: F.B., S.P.Ç., V.M.; Critical revision of the manuscript: F.B., S.P.Ç., V.M., M.N.İ. All authors reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of this study.

Ethical approval

This study was approved by the Gazi University Human Research Ethics Committe (Date: October 10, 2022, Decision/Protocol No: E-77082166). Informed consent was obtained from all participants involved in this study.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The authors declare that this study received no funding.

Generative AI statement

The authors declare that during the preparation of this study, the following AI-assisted technology was used: ChatGPT and DeepL translate on between August and September 2025. Extent of Use: for translation assistance. The authors confirm that they have critically reviewed and edited any AI-generated content and take full responsibility for the integrity, accuracy, and originality of the publication. The authors certify that the original human contribution is maintained and that AI-assisted tools are not listed or cited as authors.

Acknowledgment

We would like to thank Abdulsamet Sandal, Assoc Prof. MD, for his assistance in data collection.

References

[1] National Institute for Occupational Safety and Health (NIOSH). Criteria for a recommended standard: occupational noise exposure. Cincinnati, OH: U.S. Department of Health and Human Services; 1998.

[2] Occupational Safety and Health Administration (OSHA). Occupational noise exposure standard. 29 CFR 1910.95.

[3] Kirchner DB, Evenson E, et al. Occupational noise-induced hearing loss: ACOEM Task Force on Occupational Hearing Loss. J Occup Environ Med 2012;54(1):106-8. **[Crossref]**

[4] Nylén P, Hagman M, Johnson AC. Function of the auditory and visual systems, and of peripheral nerve, in rats after long-term combined exposure to n-hexane and methylated benzene derivatives. I. Toluene. Pharmacol Toxicol 1994;74(2):116-23. **[Crossref]**

[5] Rebert CS, Schwartz RW, Svendsgaard DJ, Pryor GT, Boyes WK. Combined effects of paired solvents on the rat's auditory system. Toxicology 1995;105(2-3):345-54. **[Crossref]**

[6] Lasky RE, Maier MM, Snodgrass EB, Hecox KE, Laughlin NK. The effects of lead on otoacoustic emissions and auditory evoked potentials in monkeys. Neurotoxicol Teratol 1995;17(6):633-44. **[Crossref]**

[7] Young JS, Upchurch MB, Kaufman MJ, Fechter LD. Carbon monoxide exposure potentiates high-frequency auditory threshold shifts induced by noise. Hear Res 1987;26(1):37-43. **[Crossref]**

[8] Morata TC, Fiorini AC, Fischer FM, et al. Toluene-induced hearing loss among rotogravure printing workers. Scand J Work Environ Health 1997;23(4):289-98. **[Crossref]**

[9] Fuente A, McPherson B, Cardemil F. Xylene-induced auditory dysfunction in humans. Ear Hear 2013;34(5):651-60. **[Crossref]**

[10] Bouchard M, Mergler D, Baldwin ME, Panisset M. Manganese cumulative exposure and symptoms: a follow-up study of alloy workers. Neurotoxicology 2008;29(4):577-83. **[Crossref]**

[11] Park RM, Bowler RM, Eggerth DE, et al. Issues in neurological risk assessment for occupational exposures: the Bay Bridge welders. Neurotoxicology 2006;27(3):373-84. **[Crossref]**

[12] Ma C, Schneider SN, Miller M, et al. Manganese accumulation in the mouse ear following systemic exposure. J Biochem Mol Toxicol 2008;22(5):305-10. **[Crossref]**

[13] SGK İstatistik Yıllıkları, Social Security Instution Year Book. 2021. Available at: https://www.sgk.gov.tr/Istatistik/Yillik/fcd5e59b-6af9-4d90-a451-ee7500eb1cb4 (Accessed on Aug, 2025).

[14] Tunay M, Melemez K. Noise induced hearing loss of forest workers in Turkey. Pak J Biol Sci 2008;11(17):2144-8. **[Crossref]**

[15] Kesici GG, Ünlü İ, Topçu AB, Bal CD, Tutkun E, Yılmaz ÖH. Arsenic related hearing loss in miners. Am J Otolaryngol 2016;37(1):6-11. **[Crossref]**

[16] Yilmaz N, Ila K, Soylemez E, Ozdek A. Evaluation of vestibular system with vHIT in industrial workers with noise-induced hearing loss. Eur Arch Otorhinolaryngol 2018;275(11):2659-65. **[Crossref]**

[17] Budak B, Çoban K, Erbek SS. Evaluatıon of the hearıng status ın carpenters. Int Arch Occup Environ Health 2021;94(7):1703-07. **[Crossref]**

[18] Sriopas A, Chapman RS, Sutammasa S, Siriwong W. Occupational noise-induced hearing loss in auto part factory workers in welding units in Thailand. J Occup Health 2017;59(1):55-62. **[Crossref]**

[19] Klockhoff I, Drettner B, Svedberg A. Computerized classification of the results of screening audiometry in groups of persons exposed to noise. Audiology 1974;13(4):326-34. **[Crossref]**

[20] World Health Organization (WHO). Physical status: the use and interpretation of anthropometry. WHO Technical Report Series No. 854. Geneva: WHO; 1995.

[21] Vermiglio AJ, Griffin S, Post C, Fang X. An evaluation of the World Health Organization and American Medical Association ratings of hearing impairment and simulated single-sided deafness. J Am Acad Audiol 2018;29(7):634-47. **[Crossref]**

[22] Delecrode CR, de Freitas TD, Frizzo ACF, Cardoso ACV. Prevalence of tinnitus in workers exposed to noise and organophosphates. Int Arch Otorhinolaryngol 2012;16(3):328-34. **[Crossref]**

[23] Kakavandi MG, Omidi A, Hashemian AH, et al. An assessment of noise exposure and hearing health status among auto body workers in Kermanshah, Iran. J Educ Health Promot 2021;10:290. **[Crossref]**

[24] Luo L, Jiang J, Huang SL, He J, Li JM. Analysis on characteristics of hearing loss in occupational noise-exposed workers in automotive manufacturing industry. Zhonghua Lao Dong Wei Sheng Zhi Ye Bing Za Zhi 2018;36(6):445-8. **[Crossref]**

[25] Gopinath B, McMahon C, Tang D, Burlutsky G, Mitchell P. Workplace noise exposure and the prevalence and 10-year incidence of age-related hearing loss. PLoS One 2021;16(7):e0255356. **[Crossref]**

[26] Le TN, Straatman LV, Lea J, Westerberg B. Current insights in noise-induced hearing loss: a literature review of the underlying mechanism, pathophysiology, asymmetry, and management options. J Otolaryngol Head Neck Surg 2017;46(1):41. **[Crossref]**

[27] Hu H, Tomita K, Kuwahara K, et al. Obesity and risk of hearing loss: a prospective cohort study. Clin Nutr 2020;39(3):870-5. **[Crossref]**

[28] Yang JR, Hidayat K, Chen CL, Li YH, Xu JY, Qin LQ. Body mass index, waist circumference, and risk of hearing loss: a meta-analysis and systematic review of observational study. Environ Health Prev Med 2020;25(1):25. **[Crossref]**

[29] Barrenäs ML, Jonsson B, Tuvemo T, Hellström PA, Lundgren M. High risk of sensorineural hearing loss in men born small for gestational age with and without obesity or height catch-up growth: a prospective longitudinal register study on birth size in 245,000 Swedish conscripts. J Clin Endocrinol Metab 2005;90(8):4452-6. **[Crossref]**

[30] Poirrier AL, Pincemail J, Van Den Ackerveken P, Lefebvre PP, Malgrange B. Oxidative stress in the cochlea: an update. Curr Med Chem 2010;17(30):3591-604. **[Crossref]**

[31] Hwang JH, Wu CC, Hsu CJ, Liu TC, Yang WS. Association of central obesity with the severity and audiometric configurations of age-related hearing impairment. Obesity (Silver Spring) 2009;17(9):1796-801. **[Crossref]**

[32] Schaffer JE. Lipotoxicity: when tissues overeat. Curr Opin Lipidol 2003;14(3):281-7. **[Crossref]**

[33] Ikeda M, Ohtsuji H. Hippuric acid, phenol, and trichloroacetic acid levels in the urine of Japanese subjects with no known exposure to organic solvents. Br J Ind Med 1969;26(2):162-4. **[Crossref]**

[34] Schäper M, Demes P, Zupanic M, Blaszkewicz M, Seeber A. Occupational toluene exposure and auditory function: results from a follow-up study. Ann Occup Hyg 2003;47(6):493-502. **[Crossref]**

[35] Sliwinska-Kowalska M, Zamyslowska-Szmytke E, Szymczak W, et al. Hearing loss among workers exposed to moderate concentrations of solvents. Scand J Work Environ Health 2001;27(5):335-42. **[Crossref]**

[36] Szulc-Kuberska J, Tronczynska J, Latkowski B. Oto-neurological investigations of chronic trichloroethylene poisoning. Minerva Otorhinolaryngol 1976;26:108-12.

[37] Muijser H, Lammers JH, Kullig BM. Effects of exposure to trichloroethylene and noise on hearing in rats. Noise Health 2000;2(6):57-66.

acta medica

RESEARCH ARTICLE

Cam-type femoro-acetabular impingement: Mid-term functional results and joint awareness of arthroscopic, mini-open and surgical dislocation techniques

Rıza Mert Çetik[1], Sancar Bakırcıoğlu[1], Kadir Büyükdoğan[2], Ömür Çağlar[1], Özgür Ahmet Atay[1], Bülent Atilla[1]

[1]Department of Orthopaedics and Traumatology, Faculty of Medicine, Hacettepe University, Ankara, Türkiye
[2]Department of Orthopedic Surgery, Güven Hospital, Ankara, Türkiye

Abstract

Objective: To evaluate the mid-term results of three different surgical techniques utilized for the treatment of femoro-acetabular impingement, by means of functional outcomes and joint awareness.

Materials and Methods: 53 patients were selected retrospectively, treated by one of the three surgical techniques: hip arthroscopy (HA), anterolateral mini open (AMO) or surgical hip dislocation (SHD). Patients with <12 months of follow-up were excluded. Median follow-up period was 57.5 months (range 16-256 months). Primary outcome measures were the Forgotten Joint Score (FJS), Harris Hip Score (HHS) and UCLA activity index. Surgical success was determined as FJS≥70 and a regression analysis was performed on different radiological and clinical parameters for the risk of failure.

Results: Mean post-operative FJS was 74.5 (±13.8) for the HA group, 72.6 (±17) for the AMO group and 52.7 (±24) for the SHD group (p=0.023). Pre-operative and post-operative HHS and UCLA were similar. Post-operative alpha angles were similar (p=0.597). Regression analysis results showed two factors affecting surgical success: pre-operative Tönnis stage (adjusted OR±95% CI: 0.091±0.005-0.659 ; p=0.023) and type of surgery performed (HA: adjusted OR±95% CI: 11.2±1.2-50.5, p=0.029; for AMO: adjusted OR±95% CI: 7.2 ±1.1-48.8, p=0.041).

Conclusion: Despite allowing more complete assessment of the hip joint, SHD results in inferior outcomes when evaluated with FJS.Three surgical approaches provide similar radiological and functional results if evaluated with conventional outcome scores of HHS and UCLA.

Keywords: femoro-acetabular impingement, hip joint, cam, hip preservation, forgotten joint score

Introduction

Cam-type femoro-acetabular impingement (FAI) occurs due to the structural deformity of the hip joint and is one of the most common causes of hip pain in young active patients [1,2]. If left untreated, it causes osteoarthritis and hip joint dysfunction in early period with accompanying cartilage and labral pathologies, leading to a decrease in the quality of life[3]. The main purpose of preventive hip surgeries applied in cam-type FAI is correction of the bony deformity, prevention of impingement and treatment of accompanying intra-articular pathologies

Corresponding author: Rıza Mert Çetik • **Email:** rmcetik@gmail.com

Received: October 21, 2025 **Accepted:** March 19, 2026 **Published online:** June 28, 2026

and improve the quality of daily life. Yet, there is still some debate on the surgical methods for the treatment of this pathology. The safe surgical dislocation defined by Ganz et al.[4] in 2001 was previously described as the gold standard, and the short- to mid-term results for this procedure are also favourable for many authors. However, Hip arthroscopy (HA) and anterolateral mini open (AMO) techniques are more preferred today because surgical hip dislocation (SHD) is a technically challenging procedure which requires ligamentum teres detachment, trochanteric osteotomy and refixation, and is more prone to intraarticular adhesions.

HA and AMO can be applied without the need of trochanteric osteotomy and sacrificing of the ligamentum teres, compared to surgical dislocation, also provide faster rehabilitation and recovery[5,6]. Although these procedures are less traumatic and have low complication rates, there are several associated drawbacks such as risk of femoral nerve palsy, inability to display entire labral or chondral lesion and the difficulty of performing osteoplasty at an appropriate amount.

Despite the existence of significant overlap of their indications, there are few studies comparing the results of these 3 surgical methods in the literature[7,8]. Although surgical dislocation was the most preferred method in our clinic in the past, in recent years HA and AMO (Figure 1) approaches have been preferred regarding shorter rehabilitation period and avoidance of implant removal. Besides, SHD is still preferred in patients presenting with global pincer and reserved for the cases that require close evaluation and manipulation of the posterior part of the hip joint.

Functional outcomes of FAI surgery were generally evaluated in the literature with clinical scoring systems such as the HSS and UCLA activity index. These scores mainly focus on the functional status and give less information about patient satisfaction.

The aim of the study is to analyse the midterm results of 3 different surgical approaches by evaluating patient satisfaction and joint awareness with a current patient reported outcome measure (PROM): Forgotten Joint Score (FJS-12).

Our hypotheses are;

1. All procedures have similar functional and radiological success without demonstrating a superiority of either technique
2. Minimally invasive procedures (HA and AMO) provides higher scores in FJS-12 than SHD
3. Pre-operative radiographic or clinical parameters indicating advanced disease adversely affect patient outcome.

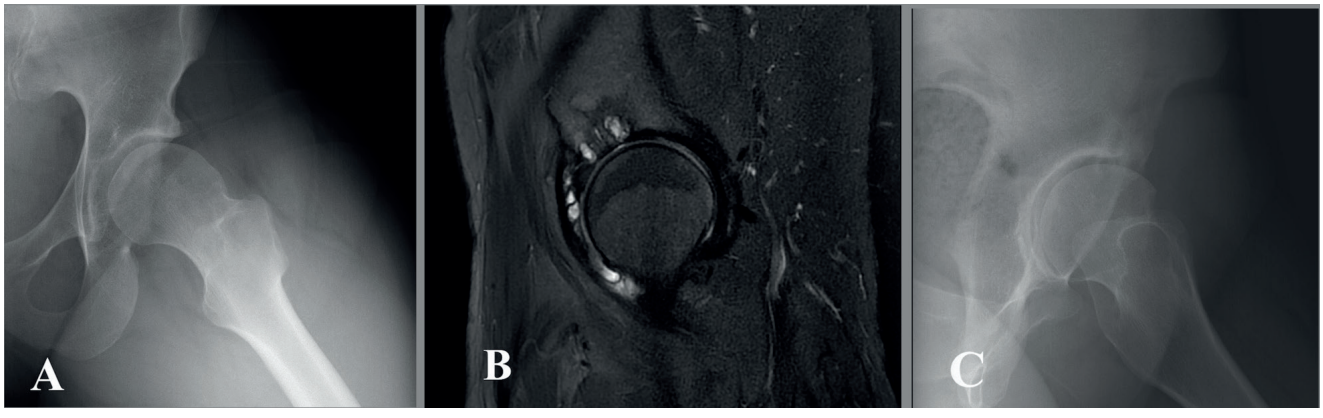


Figure 1. A 38 year-old female patient who underwent surgery with the AMO technique; a: pre-operative x-ray demonstrating an alpha angle of 89 degrees, b: a sagittal T2-weighted MRI section showing the degenerative para-labral cyst, and c: the post-operative x-ray demonstrating the resected cam lesion.

Materials and methods

Patients and study design

A single-center consecutive case study is conducted retrospectively at the Hacettepe University Hospital, Department of Orthopaedics and Traumatology.. Approval from the institutional ethical committee has been obtained. Informed consent was obtained from all patients prior to surgery. Institutional database search revealed 53 patients undergoing surgical treatment for FAI between January 1999- January 2019. Patients with <12 months of follow-up were excluded. After exclusion, 41 patients: 24 men and 17 women, with an average age of 35.8 (range 16-62) have been included in our study. The median follow-up period was 57.5 months (range 16-256 months). For further analysis, the patients are separated into three groups according to the surgical method used: HA (n=14), AMO (n=18), and SHD (n=9). The demographic and clinical information of the groups are detailed in Table 1.

Table 1. Demographic information, clinical parameters and functional status of the three study groups

	Hip Arthroscopy (HA)	Anterolateral Mini Open (AMO)	Surgical Hip Dislocation (SHD)	p value
Number of patients	14	18	9	
Age (Mean±SD)	34.5 (±12.2)	37.2 (±13.9)	35.0 (±10)	=0.874
Sex (Male/Female)	9/5	11/7	4/5	=0.655
Body mass index (Mean±SD)	24.2(±4.2)	24.9 (±3.5)	24.0 (±2.6)	=0.860
Pre-operative alpha angle (Mean±SD)	64.1 (±12.7)	67.5 (±6.6)	79.7 (±15.1)	=0.042
Follow-up period (Months; mean±SD)	60.7 (±22.4)	64.1(±53.4)	81.9 (±63.3)	=0.677
Tönnis stage				=0.560
<2	11 (78.6%)	16 (88.9%)	7 (77.8%)	
≥2	3 (21.4%)	2 (11.1%)	2 (22.2%)	
Anterior offset (Mean±SD)	0.17 (±0.03)	0.17 (±0.05)	0.16 (±0.05)	=0.440
Lateral center-edge angle (Mean±SD)	28.3 (±4.6)	26.9 (±8.6)	31.2 (±8.3)	=0.279
Acetabular angle (Mean±SD)	39.1 (±4.4)	39.7 (±4.8)	38.4 (±2.9)	=0.641
Tönnis angle (Mean±SD)	5.5 (±3.8)	5.1 (±3.8)	5.8 (±4.6)	=0.965
Labral tear				=0.492
Not present (Full thickness or degenerative)	5 (35.7%)	3 (16.7%)	3 (33.3%)	
Present	9 (64.3%)	15 (83.3%)	6 (66.7%)	
Pre-operative HHS (Mean±SD)	68.8 (±6.9)	70.5 (±8)	72.4 (±8.6)	=0.555
Post-operative HHS (Mean±SD)	88.0 (±4.9)	86.6 (±6.4)	84.3 (±9.5)	=0.579
Pre-operative UCLA (Mean±SD)	5.1 (±0.9)	4.9 (±1.2)	5.4 (±1.8)	=0.405
Post-operative UCLA (Mean±SD)	6.2 (±1.1)	6.1 (±1.2)	5.9 (±1.1)	=0.848
Post-operative alpha angle (Mean±SD)	41.9 (±11.7)	39.1 (±6.5)	37.7 (±11.9)	=0.739
Post-operative FJS (Mean±SD)	74.5 (±13.8)	72.9 (±17)	52.7 (±24)	=0.024
Surgical success according to FJS				=0.029
Successful (≥70)	11 (78.6%)	15 (83.3%)	3 (33.3%)	
Unsuccessful (<70)	3 (21.4%)	3 (16.7%)	6 (66.7%)	

Treatment

Before the surgical intervention, all patients underwent a thorough physical examination focused for hip impingement. Standard anteroposterior (AP), cross-table lateral and 45° Dunn radiographs[9], and magnetic resonance imaging (MRI) were also obtained. The alpha angle was measured on the 45° Dunn view by two of the authors (RMC, SB) Femoral head-neck offset, lateral center-edge angle (LCEA) , acetabular angle of Sharp[10] and the Tönnis angle were also measured. The extent of degeneration was evaluated according to the Tönnis classification[11]. For further analysis, the Tönnis stage was dichotomized as <2 and ≥2. MRI was examined for labral tears. As for functional assessment, each patient is asked to complete the Harris Hip Score[12] (HHS) and the UCLA activity scale [13].

Same surgical team (OAA, BA) performed all the procedures. Hip arthroscopy was performed with the patient in supine position, on a traction table. Most commonly used portals were anterior, anterolateral and posterolateral. Anterolateral mini-open surgery was performed with the patient lying supine according to Hueter's technique[14], the femoral head was exposed of anteriorly, and the cam lesion was addressed.

Post-operatively, follow-up visits were scheduled at 6 weeks, 3 months, 6 months, 12 months and yearly thereafter. Post-operative alpha angle is measured at the immediate post-operative x-ray. The HHS and UCLA activity indexes and FJS-12 were completed by the patients at the latest follow-up visits [15].

Outcomes

The primary outcome parameter was the post-operative FJS. In addition to the scores, the patients were dichotomized according to the success of treatment based on the obtained FJS. The threshold for this has been determined as 70%, which is based on clinical judgement which is also consistent with the literature regarding hip procedures[16,17]. The independent variables tested for a relationship with post-operative FJS were as follows: age, body mass index (BMI), labral tear, Tönnis stage (<2 and ≥2), pre-operative alpha angle, LCEA, acetabular angle, Tönnis angle and the type of surgery performed (HA, AMO or SHD).

The secondary outcome measures were the post-operative HHS, UCLA activity index and alpha angles.

Statistical analysis

Statistical analysis was performed using the software package SPSS (IBM Corp. Released 2015. IBM SPSS Statistics for Mac OS, Version 23.0. Armonk, NY). Descriptive statistics are presented as means-standard deviations (SD) for parametric variables and medians-ranges for non-parametric variables. When comparing means, one-way ANOVA and student's t-test were used for parametric variables; and Mann-Whitney-U test and Kruskal-Wallis test were used for non-parametric variables. Paired samples t-test was used for comparing the pre-operative and post-operative functional scores. The chi-square test and Fisher's exact test were used for comparing categorical variables between groups. We also ran a logistic regression analysis with treatment success (based on FJS) as the dependent variable. The results of the multivariate logistic regression are presented as odds ratio (OR) with 95% confidence intervals (CI). Level of significance was determined as 5% for all the tests.

Results

Pre-operative

The number of patients in the study groups and the pre-operative clinical and radiographic results are summarized in Table 1. When three groups are compared, the only variable found to be different was the pre-operative alpha angle (p=0.042). When compared in pairs, SHD group had a higher average pre-operative alpha angle than both the HA (p=0.021) and the AMO (p=0.047) groups.

Pre-operative mean HHS and UCLA activity index values (Table 1) were similar between groups (p=0.719 and p=0.405, respectively).

Post-operative

As the primary outcome measure, the mean post-operative FJS was 74.5 (±13.8) for the HA group, 72.6 (±17) for the AMO group and 52.7 (±24) for the SHD group (Figure 2). There was a significant difference between the groups (p=0.023) and when compared pairwise, SHD group was found to be significantly lower than both the HA group (p=0.012) and the AMO group (p=0.019).

When grouped according to treatment success based on the FJS, 28 patients (68.3%) resulted in a successful treatment meanwhile 13 patients (31.7%) were less successful. When success rates of the study groups are compared, a significant difference was found (p=0.039, Table 1). Table 2 summarizes the clinical and radiographic variables between the two groups. Two of the pre-operative features demonstrated significant difference between groups; the group regarded as unsuccessful had a higher percentage of patients with pre-operative Tönnis stage ≥2 (38.5% vs 7.2%, p=0.024), and also a higher pre-operative mean alpha angle (74.3 vs 66.6, p=0.042).

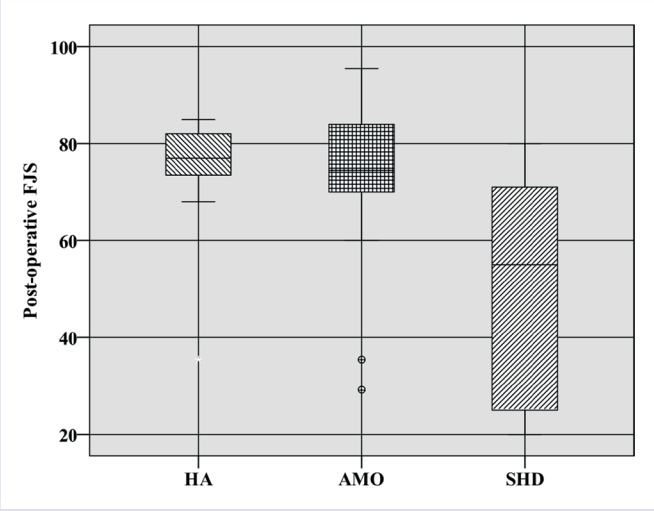


Figure 2. Post-operative FJS of the three study groups. There was a significant difference of means (p=0.024).

As the secondary outcome measures, post-operative HHS and UCLA activity indexes of the three surgery groups were compared (Table 1). The mean post-operative HHS (p=0.384) and UCLA activity indexes (p=0.848) were similar between the groups (Figure 3, Figure 4). HHS showed significant improvement after surgery, for each group, when compared to the pre-operative values: HA group increased from 68.8 (±6.9) to 88.0 (±4.9) (p<0.001), AMO group from 70.5 (±8) to 86.6 (±6.4) (p<0.001), and the SHD group from 72.4 (±8.6) to 84.3 (±9.5) (p=0.021). UCLA activity indexes also showed significant improvement for HA and AMO groups: HA group increased from 5.1 (±0.9) to 6.2 (±1.1) (p=0.003) and AMO group from 4.9 (±1.2) to 6.1 (±1.2) (p<0.001) post-operatively. The SHD group also showed a small improvement and increased from 5.4 (±1.8) to 5.9 (±1.1); but that did not reach significance statistically (p=0.225). As another secondary outcome of our study, post-operative alpha angles were also measured (Table 1). The mean values were similar between the three study groups (p=0.597). All the groups showed significant reduction of the alpha angles when compared to the pre-operative mean values: HA group decreased from 64.1 (±12.7) to 41.9 (±11.7) (p<0.001), AMO group decreased from 67.5 (±6.6) to 39.1 (±6.5) (p<0.001) and

Table 2. Clinical and radiographic variables of two groups according to treatment success based on FJS

	Unsuccessful (FJS<70)	Successful (FJS<70)	p value
Age (Mean±SD)	37.7 (±13.7)	36.4 (±11.9)	=0.699
BMI (Mean±SD)	25.2 (±2.4)	24.2 (±3.9)	=0.093
Labral tear			
Not present	2 (15.4%)	9 (32.1%)	=0.453
Present (Full thickness or degenerative)	11 (84.6%)	19 (67.9%)	
Tönnis stage			=0.024
<2	8 (61.5%)	26 (92.8%)	
≥2	5 (38.5%)	2 (7.2%)	
Pre-operative alpha angle (Mean±SD)	74.3 (±15.2)	66.6 (±9.7)	=0.106
Anterior offset (Mean±SD)	0.16 (±0.05)	0.17 (±0.04)	=0.395
Lateral center edge angle (Mean±SD)	30.8 (±8.1)	27.7 (±7.3)	=0.171
Acetabular angle (Mean±SD)	38.6 (±3.7)	39.9 (±4.5)	=0.386
Tönnis angle (Mean±SD)	4.3 (±3)	5.9 (±4.2)	=0.289

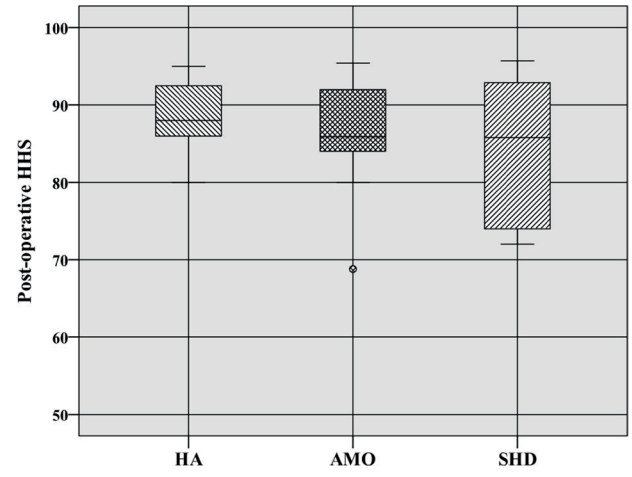


Figure 3. Post-operative HHS of the study groups. The means were not different between groups (p=0.404).

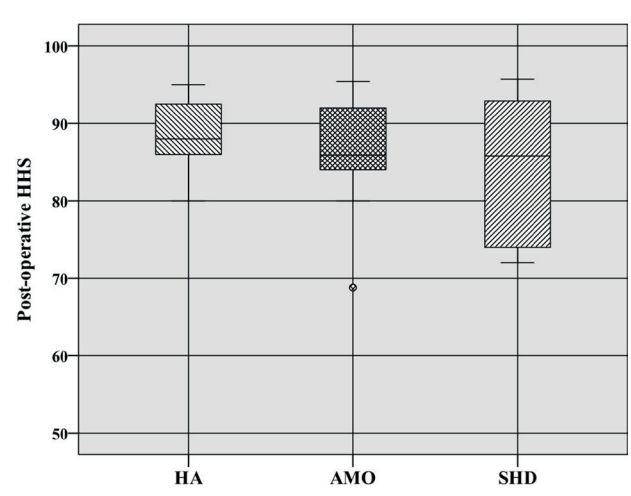


Figure 4. Post-operative UCLA activity indexes of the study groups. The means were not different between groups (p=0.848).

the SHD group decreased from 79.7 (±15.1) to 37.7 (±11.9) (p<0.001).

A multivariate logistic regression analysis was performed with treatment success as the dependent variable, to analyse the effect of pre-operative factors on the possible outcome of the treatment. The pre-operative variables that are included in the analysis were; age, BMI, presence of a labral tear, Tönnis stage (dichotomized as ≥2 and <2), pre-operative alpha angle, femoral head-neck offset, lateral CEA, acetabular angle, Tönnis angle and the type of surgery performed (with SHD designated as the baseline). The results of the regression analysis are presented in Table 3. In the univariate analysis, three of the variables (Tönnis stage, pre-operative alpha angle, type of surgery performed) reached significance and entered the multivariate analysis. In the multivariate analysis, pre-operative alpha angle lost significance and two variables remained related with treatment success: Tönnis stage and type of surgery performed.

Discussion

Even though SHD allows more complete assessment of the hip joint, labral and acetabular pathologies, the trend in recent years is in favour of less invasive techniques. Today, SHD is preferred in selected cases with global pincer, where the posterior compartment of the hip joint should be seen, while HA and AMO are frequently preferred because of both decreased hospital stay and the shorter rehabilitation period[18].

In our study, demographic data in 3 different groups were similar. Only the alpha angle was higher in the SHD group pre-operatively. This difference in the pre-operative alpha angle was an indication to go for SHD

Table 3. Results of the regression analysis

	Univariate analysis		Multivariate analysis	
	Odds ratio (95% CI)	p value	Adjusted odds ratio (95% CI)	p value
Tönnis stage	0.12 (0.02-0.74)	0.022	0.091 (±0.005-0.659)	0.023
Type of surgery				
HA	9(±1.1-71)	0.037	11.2 (±1.2-50.5)	0.029
AMO	7 (±1.2-41.4)	0.032	7.2 (±1.1-48.8)	0.041
Pre-operative alpha angle	0.94 (0.88-1.00)	0.063	-	-

in patients with severe deformity; but especially in the last few years, minimally invasive methods are more frequently utilized even for more severe deformities. Literature showed inferior results of hip arthroscopy in patients presenting with anterior acetabular insufficiency, acetabular dysplasia or aberrant femoral retroversion[19]. Although the pre-operative alpha score was higher in the SHD group, there was no statistically significant difference among the 3 groups in terms of post-operative alpha angles.

To the best of our knowledge, FJS-12[20], which is a PROM with a lower ceiling effect, was applied in 3 different surgical method groups for the first time. Our results showed that; FJS-12 was found to be lower in the SHD group, which showed comparable functional results in terms of HHS and UCLA activity indexes. Capsulotomy, trochanteric osteotomy, existence of hardware, muscle damage and sacrificing the ligamentum teres all may contribute to a micro-instability and eventually increased joint awareness in the SHD group may be the cause of difference between groups. FJS-12, which has been used since 2012 to evaluate patient satisfaction especially after arthroplasty, is needed to be analysed on FAS patients with larger cohort series.

The number of studies comparing SHD, HA and AMO is limited in the literature and also outcome measurements differ considerably among studies. In the study of Beck et al., SHD was applied to 19 hips due to FAI, and 13 patients were classified as excellent-to-good based on Merle d'Aubigné-Postel score [21]. In the study by Domb. et al., which 10 SHD and 20 HA patients were evaluated, Non-Arthritic Hip Score (NASH) was significantly higher in the HA group (94.2 vs 85.7) [22]. Additionally, in the study by Byrd et al., it is stated that during the 10-year follow-up period of hip arthroscopy performed on 15 athletes, the median improvement in the modified Harris hip score was 45 points. Also, it was emphasized that HSS is not sufficiently sensitive and has a high ceiling effect [23].

There are some limitations to our study; the mean age was relatively young compared to the literature, and there was no THA conversion in any of the patients, since the follow-up period was midterm. For this reason, joint survival was not designated as an outcome as it is in many hip-preservation studies in the literature. In the multivariate analysis, a significant relationship was found between the Tönnis stage and surgical success. In previous studies, the relationship between the increase in the Tönnis stage and the failure of surgery and the return to THA have been shown[24,25]. Risk factors for treatment failure includes a longer mean symptomatic period before surgical procedure, older age, higher BMI, hip dysplasia, acetabular retroversion, higher pre-operative alpha angle, full thickness acetabular chondral lesion, higher Tönnis grade of osteoarthritis, labral hypertrophy and total labral resection. However, its relationship on joint awareness has not been documented in the literature. The follow-up period was found to be similar between the study groups, however, the wide range of follow-up for the whole study population (ranging between 16 to 256 months) is a significant source of heterogeneity. Finally, our study has a small sample size which may not be adequate to detect small differences in the outcomes. A significant difference has been found with regards to the primary outcome measure, postoperative FJS, therefore we elected not to run a post-hoc power analysis as part of the study. However, when an analysis is made with a small effect size (0.2), the post-hoc power is found to be less than 0.5, which shows that the study may be inadequately powered to show smaller differences in other variables.

In conclusion, all 3 surgical approaches provide similar radiological and functional results after midterm follow-up if evaluated with conventional outcome scores of HHS of UCLA. In that point, FJS appears to be a more sensitive outcome measure providing accuracy to demonstrate a functional difference. We believe that joint awareness is an important component of patient satisfaction, and it is closely related to the success of a surgical procedure. High joint awareness and low patient satisfaction are expected in patients with advanced arthritis (Tönnis stage≥2) or who were applied SHD regardless of the Tönnis stage, therefore early intervention before the development of advanced arthritis and less invasive techniques (HA and AMO) are preferable for FAI surgery.

Author contributions

Conception: B.A.; Design: Ö.Ç., Ö.A.A., B.A.; Data acquisition: R.M.Ç., S.B.; Data analysis: R.M.Ç.; Data interpretation: R.M.Ç., S.B., K.B., Ö.Ç., Ö.A.A., B.A.; Drafting of the manuscript: R.M.Ç., S.B., Ö.Ç.; Critical revision of the manuscript: K.B.; All authors reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of this study.

Ethical approval

This study was approved by the Hacettepe Üniversitesi Klinik Araştırmalar Etik Kurulu (Date: January 15, 2019, Decision/Protocol No: GO209/01-15). Informed consent was obtained from all participants involved in this study.

Data availability statement

The data supporting the findings of this study are not publicly available due to containing information that could compromise the privacy of research participants.

Conflict of interest

The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The authors declare that this study received no funding.

Generative AI statement

The authors declare that no generative AI or AI-assisted technologies were used in the writing or preparation of this study.

References

[1] Ganz R, Leunig M, Leunig-Ganz K, Harris WH. The etiology of osteoarthritis of the hip: an integrated mechanical concept. Clin Orthop Relat Res 2008;466(2):264-72. **[Crossref]**

[2] Leunig M, Beaulé PE, Ganz R. The concept of femoroacetabular impingement: current status and future perspectives. Clin Orthop Relat Res 2009;467(3):616-22. **[Crossref]**

[3] Beck M, Kalhor M, Leunig M, Ganz R. Hip morphology influences the pattern of damage to the acetabular cartilage: femoroacetabular impingement as a cause of early osteoarthritis of the hip. J Bone Joint Surg Br 2005;87(7):1012-8. **[Crossref]**

[4] Ganz R, Gill TJ, Gautier E, Ganz K, Krügel N, Berlemann U. Surgical dislocation of the adult hip a technique with full access to the femoral head and acetabulum without the risk of avascular necrosis. J Bone Joint Surg Br 2001;83(8):1119-24. **[Crossref]**

[5] Bellotti V, Cardenas C, Astarita E, et al. Mini-open approach for femoroacetabular impingement: 10 years experience and evolved indications. Hip Int. 2016;26 (1 Suppl):38-42. **[Crossref]**

[6] Larson CM, Giveans MR. Arthroscopic management of femoroacetabular impingement: early outcomes measures. Arthroscopy 2008;24(5):540-6. **[Crossref]**

[7] Matsuda DK, Carlisle JC, Arthurs SC, Wierks CH, Philippon MJ. Comparative systematic review of the open dislocation, mini-open, and arthroscopic surgeries for femoroacetabular impingement. Arthroscopy 2011;27(2):252-69. **[Crossref]**

[8] Rego PA, Mascarenhas V, Oliveira FS, Pinto PC, Sampaio E, Monteiro J. Arthroscopic versus open treatment of cam-type femoro-acetabular impingement: retrospective cohort clinical study. Int Orthop 2018;42(4):791-7. **[Crossref]**

[9] Meyer DC, Beck M, Ellis T, Ganz R, Leunig M. Comparison of six radiographic projections to assess femoral head/neck asphericity. Clin Orthop Relat Res 2006;445:181-5. **[Crossref]**

[10] Sharp I. Acetabular dysplasia. The acetabular angle. J Bone Joint Surg. 1961;43(2):268-72. **[Crossref]**

[11] Tönnis D, Heinecke A. Acetabular and femoral anteversion: relationship with osteoarthritis of the hip. J Bone Joint Surg Am 1999;81(12):1747-70. **[Crossref]**

[12] Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. J Bone Joint Surg Am 1969;51(4):737-55. **[Crossref]**

[13] Zahiri CA, Schmalzried TP, Szuszczewicz ES, Amstutz HC. Assessing activity in joint replacement patients. J Arthroplasty 1998;13(8):890-5. **[Crossref]**

[14] Hueter C. Grundriss der chirurgie. In: Fünfte abtheilung: die verletzung und krankheiten des hüftgelenkes, neunundzwanzigstes capitel. 2nd ed. Leipzig: FCW Vogel;1883:129-200.

[15] Behrend H, Giesinger K, Giesinger JM, Kuster MS. The "forgotten joint" as the ultimate goal in joint arthroplasty: validation of a new patient-reported outcome measure. J Arthroplasty 2012;27(3):430-436.e1. **[Crossref]**

[16] Galea VP, Ingelsrud LH, Florissi I, et al. Patient-acceptable symptom state for the Oxford Hip Score and Forgotten Joint Score at 3 months, 1 year, and 2 years following total hip arthroplasty: a registry-based study of 597 cases. Acta Orthop 2020;91(4):372-7. **[Crossref]**

[17] Rosinsky PJ, Chen JW, Lall AC, Shapira J, Maldonado DR, Domb BG. Can we help patients forget their joint? Determining a threshold for successful outcome for the forgotten joint score. J Arthroplasty 2020;35(1):153-9. **[Crossref]**

[18] Zingg PO, Ulbrich EJ, Buehler TC, Kalberer F, Poutawera VR, Dora C. Surgical hip dislocation versus hip arthroscopy for femoroacetabular impingement: clinical and morphological short-term results. Arch Orthop Trauma Surg 2013;133(1):69-79. **[Crossref]**

[19] Makhni EC, Ramkumar PN, Cvetanovich G, Nho SJ. Approach to the patient with failed hip arthroscopy for labral tears and femoroacetabular impingement. J Am Acad Orthop Surg 2020;28(13):538-45. **[Crossref]**

[20] Hamilton DF, Giesinger JM, MacDonald DJ, Simpson AHRW, Howie CR, Giesinger K. Responsiveness and ceiling effects of the Forgotten Joint Score-12 following total hip arthroplasty. Bone Joint Res 2016;5(3):87-91. **[Crossref]**

[21] Beck M, Leunig M, Parvizi J, Boutier V, Wyss D, Ganz R. Anterior femoroacetabular impingement: part II. midterm results of surgical treatment. Clin Orthop Relat Res. Jan 2004;(418):67-73. **[Crossref]**

[22] Domb BG, Stake CE, Botser IB, Jackson TJ. Surgical dislocation of the hip versus arthroscopic treatment of femoroacetabular impingement: a prospective matched-pair study with average 2-year follow-up. Arthroscopy 2013;29(9):1506-13. **[Crossref]**

[23] Byrd JW, Jones KS. Hip arthroscopy in athletes: 10-year follow-up. Am J Sports Med Nov 2009;37(11):2140-3. **[Crossref]**

[24] Botser IB, Smith TW, Jr., Nasser R, Domb BG. Open surgical dislocation versus arthroscopy for femoroacetabular impingement: a comparison of clinical outcomes. Arthroscopy 2011;27(2):270-8. **[Crossref]**

[25] Nwachukwu BU, Rebolledo BJ, McCormick F, Rosas S, Harris JD, Kelly BT. Arthroscopic versus open treatment of femoroacetabular impingement: a systematic review of medium- to long-term outcomes. Am J Sports Med 2016;44(4):1062-8. **[Crossref]**

acta medica

RESEARCH ARTICLE

Predictive value of the HALP score for lymph node metastasis in resectable gastric cancer

Yasin Orhan Erkuş[1], Serhan Yılmaz[1], Canbert Çelik[1], Ali Sapmaz[1], Onur Öztel[1], Zeyneddin Ali Muhammed[1]

[1]Department of General Surgery, Ankara Bilkent City Hospital, University of Health Sciences, Ankara, Türkiye

Abstract

Objective: This study aimed to evaluate the predictive value of the hemoglobin, albumin, lymphocyte, and platelet (HALP) score for lymph node metastasis (LNM) in patients with resectable gastric cancer and to investigate the relationship between hematological markers and LNM.

Materials and Methods: Patients who underwent surgery for gastric adenocarcinoma between 2020 and 2024 were retrospectively analyzed. Demographic data, comorbidities, and laboratory parameters were recorded, and the HALP score, neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and platelet-to-neutrophil ratio (PNR) were calculated. Tumor stage, nodal stage, and pathological features were reviewed. The optimal HALP cutoff value was determined by receiver operating characteristic (ROC) curve analysis. Variables significant in univariate analysis were included in multivariate logistic regression to identify independent risk factors for LNM.

Results: A total of 238 patients were included, with a mean age of 65.5 ± 11.4 years; 70.2% were male. The mean number of dissected lymph nodes was 26.0 ± 10.6, and the mean number of metastatic nodes was 6.2 ± 9.0. Perineural invasion (PNI) and lymphovascular invasion (LVI) were observed in 66.8% and 71.0% of patients, respectively. LNM was present in 159 patients (66.8%). In univariate analysis, a low HALP score, advanced invasion depth, and the presence of PNI and LVI were significantly associated with LNM ($p < 0.005$). ROC analysis identified an optimal HALP cutoff value of 17 (AUC = 0.581, $p = 0.041$). In multivariate analysis, invasion depth ($p = 0.001$), PNI ($p = 0.008$), and LVI ($p < 0.001$) were independent predictors of LNM, whereas the HALP score was not ($p = 0.221$).

Conclusion: In resectable gastric cancer, a low HALP score was associated with lymph node metastasis but was not confirmed as an independent predictive factor. As a composite parameter reflecting systemic inflammation and immune–nutritional status, the HALP score may indicate tumor aggressiveness and poor prognosis. Preoperative recognition of low HALP(<17) scores may help identify patients requiring closer monitoring for LNM risk.

Keywords: gastric cancer, lymphatic metastasis, prognostic factors, HALP score

Introduction

Gastric cancer is the fifth most common malignancy worldwide and remains one of the leading causes of cancer-related mortality [1]. Adenocarcinoma is the predominant histopathological type, accounting for approximately 95% of all cases [2]. Owing to the fact that gastric cancer is frequently diagnosed at advanced stages, the overall prognosis is generally poor. The TNM staging system, which incorporates tumor depth, lymph node involvement, and distant metastasis, is the most widely accepted and clinically relevant determinant of prognosis [3].

Corresponding author: Yasin Orhan Erkuş • **Email:** yasinorhanerkus@gmail.com

Received: October 27, 2025 **Accepted:** March 6, 2026 **Published online:** June 28, 2026

In recent years, several hematological and biochemical parameters—including the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), C-reactive protein (CRP), and albumin-to-alkaline phosphatase ratio (AAPR)—have been investigated for their potential role in predicting lymph node metastasis. Findings suggest that these markers may provide additional prognostic information in gastric cancer [4,5]. Moreover, the hemoglobin, albumin, lymphocyte, and platelet (HALP) score, which reflects both nutritional and immunological status, has emerged as a novel biomarker with reported predictive value for lymph node metastasis [6,7]. For example, Çağlıyan et al. demonstrated that a lower HALP score was associated with lymph node metastasis and poorer prognosis in patients with colorectal cancer [8]. Other studies similarly suggest that a low HALP score correlates with advanced tumor stage, increased lymph node involvement, and decreased survival rates [9]. Nevertheless, evidence regarding the clinical utility of the HALP score in gastric cancer remains limited. Although previous studies examined HALP in predicting survival, its role in predicting LNM in the Turkish population remains unclear.

The primary aim of this study was to evaluate the predictive value of the HALP score for lymph node metastasis in patients with gastric cancer. A secondary objective was to explore the association between other hematological markers and lymph node metastasis.

Materials and methods

This study was approved by the Ethics Committee of Bilkent City Hospital (Decision No: TABED 1-25-1790) and conducted in accordance with the principles of the Declaration of Helsinki (revised in 2013). A retrospective cohort design was employed. Patients aged 18 years or older who underwent surgery for gastric adenocarcinoma at our clinic between 2020 and 2024 were included. Exclusion criteria were as follows: gastrectomy performed for indications other than adenocarcinoma (e.g., neuroendocrine tumors, lymphoma), concomitant malignancies, chronic inflammatory, hematological, or autoimmune diseases that could influence hematological parameters, history of corticosteroid use, presence of acute infection, receipt of neoadjuvant therapy (to avoid confounding of hematologic parameters), availability of fewer than 15 lymph nodes in the pathology specimen, and incomplete clinical data.

Data were retrieved from patient medical records and the hospital database. Demographic variables (age, sex), comorbidities (diabetes mellitus, hypertension, coronary artery disease, chronic obstructive pulmonary disease), and laboratory parameters [hemoglobin (g/dL), albumin (g/L), neutrophils ($\times10^9$/L), lymphocytes ($\times10^9$/L), platelets ($\times10^9$/L)] were recorded. Derived indices were calculated as follows: platelet-to-lymphocyte ratio (PLR = platelets/lymphocytes), platelet-to-neutrophil ratio (PNR = platelets/neutrophils), and HALP score [hemoglobin (g/dL) × albumin (g/L) × lymphocytes ($\times10^9$/L) / platelets ($\times10^9$/L)]. Laboratory values were obtained from blood samples collected within one week prior to surgery.

Clinicopathological features, including tumor stage (T), nodal status (N), overall stage, perineural invasion (PNI), lymphovascular invasion (LVI), total number of lymph nodes, and number of positive lymph nodes, were extracted from postoperative pathology reports. Tumor, nodal, and stage classifications were based on the 8th edition of the American Joint Committee on Cancer (AJCC) staging system [3]. For additional analysis, tumor invasion depth was categorized into two groups according to T stage: T1–2 and T3–4.

Statistical analysis

Categorical variables were presented as frequency and percentage and continuous variables as means and standard deviation. Mann-Whitney U test was used to compare continuous variables. Fisher's Exact Chi-square test was used to compare categorical variables. The ROC curve was used to determine the optimum cut-off values for HALP score which is associated with lymph node metastasis (based on Youden's J index from the ROC curve). Variables that showed significant correlation in the univariate analysis were evaluated by multivariate binary logistic regression analysis to determine the risk factors associated with lymph node metastasis and to calculate the odds ratio with 95% CI. All analyses were performed using the Statistical Package for the Social Sciences for Windows version 22.0 (SPSS Inc., Chicago, Illinois, USA). The level of statistical significance was set at $p<0.05$.

Results

A total of 238 patients were included in the study. The mean age was 65.5 ± 11.4 years, and the cohort consisted of 71 females (29.8%) and 167 males (70.2%). The mean follow-up duration was 27.4 ± 17.2 months. Total gastrectomy was performed in 136 patients (57.1%), while 102 patients (42.9%) underwent subtotal gastrectomy. The demographic characteristics of the study population are summarized in Table 1.

When clinicopathological data were evaluated, the mean number of dissected lymph nodes was 26.0 ± 10.6, while the mean number of metastatic lymph nodes was 6.2 ± 9.0. Perineural invasion (PNI) was identified in 159 patients (66.8%), and lymphovascular invasion (LVI) was present in 169 patients (71.0%). The clinicopathological characteristics of the patients are summarized in Table 2.

Table 1. Demographic data

Variable	Value
Age (years) (mean±SD)	65.53±11.45
Gender (n/%)	
Female	71 (29.8%)
Male	167 (70.2%)
Smoking Status (n/%)	
Non-smoker	141 (59.2%)
Smoker	97 (40.8%)
DM (n/%)	
Non-present	182 (76.5%)
Present	56 (23.5%)
HT (n/%)	
Non-present	134 (56.3%)
Present	104 (43.7%)
CAD (n/%)	
Non-present	174 (73.1%)
Present	64 (26.9%)
COPD (n/%)	
Non-present	223 (93.7%)
Present	15 (6.3%)
HALP score (mean±SD)	36.54±30.16
NLR (mean±SD)	3.24±3.16
PLR (mean±SD)	187.17±95.83
PNR (mean±SD)	66.19±26.55

SD: standart deviation

Table 2. Clinicopathological data

Variable	Value
T (n/%)	
1A	18 (7.6%)
1B	17 (7.1%)
2	15 (6.3%)
3	114 (47.9%)
4A	74 (31.1%)
N (n/%)	
0	79 (33.2%)
1	38 (16%)
2	45 (18.9%)
3A	46 (19.3%)
3B	30 (12.6%)
Stage (n/%)	
1A	33 (13.9%)
1B	12 (5%)
2A	32 (13.4%)
2B	31 (13%)
3A	37 (15.5%)
3B	48 (20.2%)
3C	45 (18.9%)
Perineural invasion	
None	79 (33.2%)
Yes	159 (66.8%)
Lymphovascular invasion	
None	69 (29%)
Yes	169 (71%)
Total lymph node number (mean±SD)	26±10.6
Positive lymph node number (mean±SD)	6.21±8.98
Depth of invasion(n/%)	
T1/2	50 (21%)
T3/4	188 (79%)

SD: standart deviation

Patients were classified into two groups: those without lymph node metastasis (N0) (79 patients, 33.2%) and those with lymph node metastasis (N1–3) (159 patients, 66.8%). In univariate analysis, the HALP score, tumor invasion depth, perineural invasion, and lymphovascular invasion were significantly associated with lymph node metastasis ($p < 0.005$) (Table 3).

The optimal cutoff value for the HALP score was determined to be 17, based on receiver operating characteristic (ROC) curve analysis using Youden's J

Table 3. Univariate analysis of factors associated with lymph node metastasis

	LNM (-) (n=79)	LNM (+) (n=159)	P value
Age (years) (mean±SD)	66.02±12.50	65.28±10.93	0.217 [a]
Gender (n/%)			0.548 [b]
Female	26 (32.9%)	45 (28.3%)	
Male	53 (67.1%)	114 (71.7%)	
DM (n/%)			0.194 [b]
Non-present	56 (70.9%)	126 (79.2%)	
Present	23 (29.1%)	33 (20.8%)	
HT (n/%)			1.000 [b]
Non-present	45 (57%)	89 (56%)	
Present	34 (43%)	70 (44%)	
CAD (n/%)			0.278 [b]
Non-present	54 (68.4%)	120 (75.5%)	
Present	25 (31.6%)	39 (24.5%)	
COPD (n/%)			0.779 [b]
Non-present	75 (94.9%)	148 (93.1%)	
Present	4 (5.1%)	11 (6.9%)	
HALP score (mean±SD)	38.75±22.20	35.44±33.42	**0.041** [a]
NLR (mean±SD)	3.39±4.87	3.16±1.81	0.283 [a]
PLR (mean±SD)	173.94±95.46	193.74±95.63	0.074 [a]
PNR (mean±SD)	64.53±28.11	67.02±25.79	0.360 [a]
Depth of invasion(n/%)			**<0.001** [b]
T1/2	43 (54.4%)	7 (4.4%)	
T3/4	36 (45.6%)	152 (95.6%)	
Perineural invasion			**<0.001** [b]
None	59 (74.7%)	20 (12.6%)	
Yes	20 (25.3%)	139 (87.4%)	
Lymphovascular invasion			**<0.001** [b]
None	57 (72.2%)	12 (7.5%)	
Yes	22 (27.8%)	147 (92.5%)	

SD: standart deviation, [a] Mann-Whitney U Test, [b] Fisher's Exact Test

index. This threshold yielded a sensitivity of 86.1% and a specificity of 29.6% (AUC: 0.581, 95% CI: 0.505–0.657, p = 0.041) (Figure 1).

Variables that were significant in univariate analysis were subsequently included in the multivariate logistic regression model. Tumor invasion depth (OR: 0.127, 95% CI: 0.039–0.413, p = 0.001), perineural invasion (OR: 0.276, 95% CI: 0.107–0.716, p = 0.008), and lymphovascular invasion (OR: 0.055, 95% CI: 0.021–0.140, p < 0.001) were identified as independent risk factors for lymph node metastasis (Table 4).

Discussion

In the present study, univariate analyses demonstrated a significant association between the HALP score, tumor invasion depth, perineural invasion (PNI), lymphovascular invasion (LVI), and the presence of lymph node metastasis (LNM) in patients with resectable gastric adenocarcinoma. However, in multivariate logistic regression analysis, only invasion depth, PNI, and LVI were identified as independent predictive factors for LNM, whereas the HALP score did not retain statistical significance. Receiver operating characteristic (ROC) analysis indicated a HALP cutoff value of 17, which yielded high sensitivity (86.1%) but low specificity (29.1%) (AUC: 0.581, 95% CI: 0.505–0.657, p = 0.041). PNI and LVI were observed in 66.8% and 71.0% of patients, respectively, indicating a high prevalence of aggressive pathology included in the model. This suggests that the effect of hematological parameters may be overshadowed. Also the low specificity of HALP cut-off value can be explained by this way. While this limits the utility of the HALP score as a standalone predictor of LNM, it still holds value as a complementary marker in the broader prognostic assessment of gastric cancer.

Wang et al. reported that the HALP score was an independent risk factor for LNM in patients with gastric cancer, with an odds ratio of 2.276 (p = 0.032) [10]. In that study, the HALP score, when combined with tumor invasion depth and tumor markers (CEA, CA19-9), demonstrated strong predictive performance for LNM. In our study HALP score was significantly higher in the lymph node–positive group (p = 0.041). Differences in study populations, sample sizes, and the range of variables incorporated into statistical models may partly explain these discrepancies.

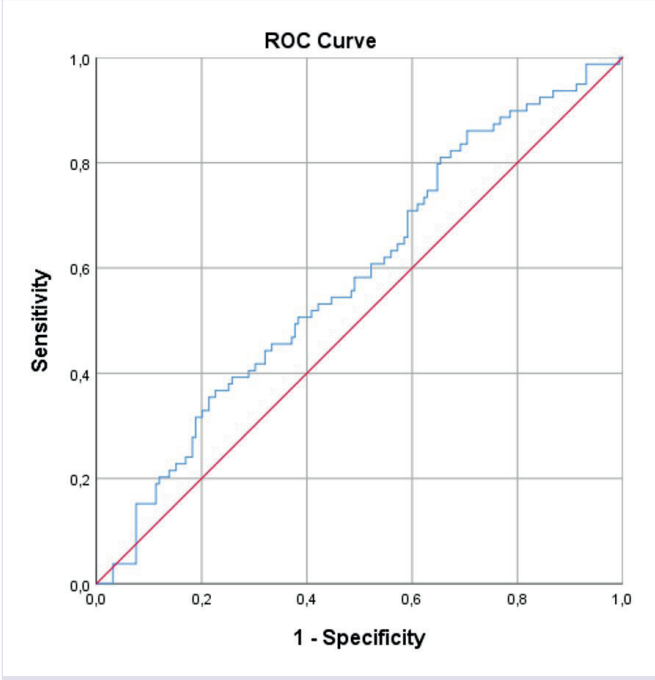


Figure 1. ROC curve for HALP score

Table 4. Multivariate analysis of factors associated with lymph node metastasis

	OR	95% CI	p value
HALP score	1.898	0.680-5.294	0.221
Depth of invasion(n/%)	0.127	0.039-0.413	**0.001**
Perineural invasion	0.276	0.107-0.716	**0.008**
Lymphovascular invasion	0.055	0.021-0.140	**<0.001**

OR: Odds ratio, CI: Confidence interval

Chen et al. analyzed a large cohort of 1,332 patients and demonstrated that the HALP score was closely associated with clinicopathological characteristics, as well as survival outcomes [11]. Patients with a HALP score <56.8 were more likely to present with advanced T and N stages, whereas those with a HALP score ≥56.8 exhibited significantly longer median survival and improved 1-, 2-, and 3-year survival rates. These results indicate that worsening prognosis parallels the increased frequency of lymph node metastasis, and that lymph node involvement adversely impacts overall survival.

In a more recent study, Aoyama et al. (2024) reported 5-year overall survival rates of 57.2% in patients with HALP ≤40 and 85.8% in those with HALP >40

who underwent curative gastrectomy. Furthermore, multivariate analysis identified a low HALP score as an independent risk factor for survival (HR = 2.679, p = 0.002). Similarly, Sargın et al. demonstrated that patients with higher HALP scores had significantly longer overall survival in gastric carcinoma [12].

Taken together, these findings suggest that a low HALP score reflects impaired nutritional and immune status, which may contribute to more aggressive tumor biology, progression, and worse survival outcomes. HALP score is statistical significance in univariate analysis—combined with robust evidence from prior studies—supports its potential role as a clinically valuable prognostic marker in gastric cancer.

Köşeci et al. evaluated the role of the HALP score in predicting perioperative treatment response in early-stage gastric cancer and reported a significantly higher pathological response rate among patients with a HALP score ≥28.9 [13]. In our study, treatment response parameters were not analyzed because neoadjuvant therapy was not administered, as such treatment could potentially influence hematological parameters. Nevertheless, the findings of Köşeci et al. support the notion that the HALP score may reflect tumor biology and sensitivity to therapy.

In our series, other hematological inflammatory indices such as the neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR), and platelet-to-neutrophil ratio (PNR) were not significantly associated with lymph node metastasis. However, numerous studies in the literature have demonstrated the prognostic relevance of NLR and PLR in gastric cancer. For example, Gunaldi et al. analyzed 245 patients in Turkey and showed that patients with elevated NLR were more frequently diagnosed at advanced stages (55.9% with high NLR) and were more likely to present with lymph node metastases. Elevated PLR was correlated with deeper tumor invasion (T3/4) and higher stage. Furthermore, this study reported that baseline NLR at the time of diagnosis was a significant prognostic factor for survival and could even aid in guiding surgical and treatment strategies. Interestingly, PLR did not show a significant association with survival in the same cohort [4].

Consistent with these findings, Zhang et al. conducted a large meta-analysis and demonstrated that elevated preoperative PLR was associated with poorer prognosis in gastric cancer patients (HR = 1.37, 95% CI: 1.26–1.49) and significantly increased the risk of lymph node metastasis (OR = 1.17, 95% CI: 1.02–1.33) [6]. Although no statistical association was identified for PLR or PNR in our cohort, the cumulative evidence suggests that systemic inflammatory markers may provide insight into tumor biology and metastatic potential. Variations in cutoff values, sample sizes, and heterogeneity among study populations likely contribute to the discrepancies observed across studies. Among these indices, the prognostic significance of NLR appears more consistent, whereas the role of PLR remains less definitive.

The findings of our study indicate that a low HALP score may be associated with more aggressive pathological features and the presence of lymph node metastasis in gastric cancer. As a composite parameter reflecting both systemic inflammation and immune–nutritional status, the HALP score provides a broader prognostic perspective than individual hematological indices. Although its specificity is limited, the HALP score should not be overlooked in clinical practice. In the preoperative setting, a low HALP score—easily obtained without additional investigations—may serve as a practical marker to identify patients who require closer monitoring and heightened vigilance regarding the risk of lymph node metastasis. Although HALP was not an independent predictor, its easy preoperative calculation may complement existing risk models. Prospective multicenter validation is warranted.

Taken together with evidence from the current literature, our results suggest that integrated indices such as the HALP score may hold potential value not only for prognostic assessment but also for guiding treatment planning in patients with gastric cancer.

Author contribution

Conception: Y.O.E., S.Y., C.Ç.; Design: Y.O.E., S.Y., C.Ç.; Data acquisition: Y.O.E., O.Ö., Z.A.M.; Data analysis: Y.O.E., S.Y., C.Ç., A.S.; Data interpretation: Y.O.E., S.Y., C.Ç., A.S.; Drafting of the manuscript: Y.O.E., S.Y.; Critical revision of the manuscript: Y.O.E., S.Y., C.Ç., A.S., O.Ö., Z.A.M. All authors reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of this study.

Ethical approval

This study was approved by the Ethics Committee of Bilkent City Hospital (Date: October 22, 2025, Decision/ Protocol No: TABED 1-25-1790). Informed consent was obtained from all participants involved in this study.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The authors declare that this study received no funding.

Generative AI statement

The authors declare that no generative AI or AI-assisted technologies were used in the writing or preparation of this study.

References

[1] Sung H, Ferlay J, Siegel RL, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2021;71(3):209-49. **[Crossref]**

[2] Ma J, Shen H, Kapesa L, Zeng S. Lauren classification and individualized chemotherapy in gastric cancer. Oncol Lett 2016;11(5):2959-64. **[Crossref]**

[3] Amin MB, Greene FL, Edge SB, et al. The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more "personalized" approach to cancer staging. CA Cancer J Clin 2017;67(2):93-9. **[Crossref]**

[4] Gunaldi M, Goksu S, Erdem D, et al. Prognostic impact of platelet/lymphocyte and neutrophil/lymphocyte ratios in patients with gastric cancer: a multicenter study. Int J Clin Exp Med 2015;8(4):5937-42.

[5] Yamanaka T, Matsumoto S, Teramukai S, Ishiwata R, Nagai Y, Fukushima M. The baseline ratio of neutrophils to lymphocytes is associated with patient prognosis in advanced gastric cancer. Oncology 2007;73(3-4):215-20. **[Crossref]**

[6] Zhang T, Liu W, Xu C. Correlation analysis of hemoglobin, albumin, lymphocyte, platelet score and platelet to albumin ratio and prognosis in patients with lung adenosquamous carcinoma. Front Oncol 2023;13:1166802. **[Crossref]**

[7] Yalav O, Topal U, Unal AG, Eray IC. Prognostic significance of preoperative hemoglobin and albumin levels and lymphocyte and platelet counts (HALP) in patients undergoing curative resection for colorectal cancer. Ann Ital Chir 2021;92:283-92.

[8] Çağlıyan Ö, Yazıcı H, Yaşar AC, et al. Could the HALP score indicate poor prognosis in colorectal cancer patients? Turk J Surg 2025;41(2):154-9. **[Crossref]**

[9] Xu SS, Li S, Xu HX, et al. Haemoglobin, albumin, lymphocyte and platelet predicts postoperative survival in pancreatic cancer. World J Gastroenterol 2020;26(8):828-38. **[Crossref]**

[10] Wang X, He Q, Liang H, et al. A novel robust nomogram based on preoperative hemoglobin and albumin levels and lymphocyte and platelet counts (HALP) for predicting lymph node metastasis of gastric cancer. J Gastrointest Oncol 2021;12(6):2706-18. **[Crossref]**

[11] Chen XL, Xue L, Wang W, et al. Prognostic significance of the combination of preoperative hemoglobin, albumin, lymphocyte and platelet in patients with gastric carcinoma: a retrospective cohort study. Oncotarget 2015;6(38):41370-82. **[Crossref]**

[12] Aoyama T, Maezawa Y, Hashimoto I, et al. The clinical impact of Hemoglobin, Albumin, Lymphocyte, Platelet (HALP) in gastric cancer patients who receive curative treatment. In Vivo 2024;38(5):2494-500. **[Crossref]**

[13] Köşeci T, Seyyar M, Aydınalp Camadan Y, et al. HALP score in predicting response to treatment in patients with early-stage gastric cancer: a multi-centred retrospective cohort study. Medicina (Kaunas) 2024;60(12):2087. **[Crossref]**

Acta Medica 2026; Early View: 139-145 DOI: 10.32552/actamedica.2026.1212

acta medica RESEARCH ARTICLE

The performance of ChatGPT and Google Bard in medical oncology board examination

Taha Koray Şahin[1], Murat Dinçer[2], Nuri Karadurmuş[3], Deniz Can Güven[1]

[1]Department of Medical Oncology, Hacettepe University, Ankara, Türkiye
[2]Department of Medical Oncology, Eskişehir Osmangazi University, Eskişehir, Türkiye
[3]Department of Medical Oncology, Gulhane School of Medicine, University of Health Sciences, Ankara, Türkiye

Abstract

Objective: Artificial intelligence (AI) is transforming healthcare, and large language models (LLMs) like ChatGPT and Google Bard have shown promise in providing medical information and decision support. The LLMs performed similarly or better than human participants in several board exams. However, their proficiency in complex clinical scenarios, like in oncology board exams, remains unclear. We aimed to assess the performance of three LLMs (ChatGPT 3.5, ChatGPT 4 and Google Bard) on the oncology board examination.

Materials and Methods: We utilized a question bank from the Turkish Society of Medical Oncology Board Exam encompassing 290 multiple-choice questions from 2021-2023. ChatGPT 3.5, ChatGPT 4, and Google Bard were asked to answer each question in both Turkish and English, providing explanations and confidence levels with their answers.

Results: The overall accuracy of LLMs was 59.3%, 42.8%, 36.2% for ChatGPT4, ChatGPT3.5, and Google Bard, respectively. The accuracy of ChatGPT 4 was significantly higher than that of ChatGPT 3.5 ($p<0.001$) and Google Bard ($p<0.001$), while the accuracy of ChatGPT3.5 was higher than that of Google Bard ($p<0.001$). Only the ChatGPT 4 was proficient in all three examination years (2021-2023). All LLMs performed better on translated questions than original Turkish ones. The LLMs were more accurate in general knowledge than case questions and were more confident in their answers for translated questions.

Conclusion: LLMs had moderate success in a medical oncology board exam, with only ChatGPT 4 demonstrating proficiency. The efficacy of LLMs in clinical decision-making requires further development, especially in native languages and complex case interpretations.

Keywords: large language models, ChatGPT, Google Bard, medical oncology, board, exam

Introduction

Artificial intelligence (AI) can transform every aspect of daily living, including healthcare [1]. The field of oncology was one of the most studied specialties in which AI emerged as a feasible way to improve patient care. In this regard, AI has already been used in cancer screening, molecular pathology, and big data analysis [2,3]. Furthermore, AI was promising as a decision-making tool after the earlier success of several models like IBM Watson [4]. However, the interest in this area was diminished after the limited clinical benefit of these models in complex clinical scenarios. Although not primarily developed for clinical decision and healthcare, the newly created large language models (LLMs) could potentially counter the limitations of previous

Corresponding author: Taha Koray Sahin • **Email:** takorsah@gmail.com

Received: November 12, 2025 **Accepted:** February 03, 2026 **Published online:** June 28, 2026

clinical decision support systems due to their training with larger datasets [5,6]. However, the potential of these LLMs in resolving complex clinical scenarios and generating true medical information has not been thoroughly investigated.

The most important prominent members of the recently developed LLMs are ChatGPT 3.5, ChatGPT 4, and Google Bard. In earlier studies, these LLMs were able to give accurate medical information in over the clinical scenarios [7,8]. Additionally, to evaluate their performance against human intelligence, these LLMs were tested in medical board exams. While the earlier results were very promising and demonstrated proficiency in board exams like United States Medical Licensing Examination (USMLE) [9] and radiology board exams [10], the LLMs' success was lower in specialties like ophthalmology [11] and neurology [12], in which more complex clinical scenarios and higher-order questions were more frequent. Additionally, the performance of individual LLMs, as well as the versions of the individual LLMs varied [11,13-15]. Despite this body of evidence in cardiology [16], radiology [17], and surgery boards [18], the performance of LLMs was not investigated in oncology board exams. Considering the huge burden of cancer, the generation of accurate information and the proficiency of LLMs in oncology is paramount. Therefore, we evaluated the performance of LLMs in an oncology board exam (Turkish Society of Medical Oncology Board Exam) and compared the performance of individual LLMs.

Methods

Sample questions

A question bank comprising the questions from the last three years (2021-2023) and including a total of 290 questions was used (Supplement). The questions for the board certification of Medical Oncologists from Turkey were created by a Turkish Society of Medical Oncology Proficiency Board Members, and the passing grades for each year were previously calculated according to the difficulty of this year's exam. The passing score was 53, 49, and 58 for 2021, 2022 and 2023, respectively. The full question set used in this study is openly available

The questions were multiple choice questions with five options and one correct answer with four distractors. The questions were retrieved from the website of the Turkish Society of Medical Oncology (members area). They were used with the approval of the Turkish Society of Medical Oncology Executive Board. The proficiency board members previously calculated the individual question difficulties, and these difficulty levels were used to compare the difficulty level annotated by the LLMs. Question topics and formats were categorized as case-based or general knowledge questions.

Data collection

The ChatGPT 3.5, ChatGPT 4, and Google Bard LLMs were used via the individual website interfaces. While the previous ChatGPT models were trained up to September 2021, this restriction was removed on September 27th, 2023, and both ChatGPT versions have access to live data via internet browsing. Similarly, the Google Bard could have live data from the internet. Therefore, all three LLMs were expected to provide answers in light of the most up-to-date data.

The researchers did not additionally pre-trained the LLMs before replying to the questions. The questions were asked in Turkish, and the LLMs were asked to answer in Turkish whenever possible. In another turn, the LLMs were also asked to translate the questions into English and give answers in English with explanations.

The following command was given to individual LLMs to gather data and all answer choices to this command and explanations were recorded.

"You are a medical oncologist and you are taking the oncology board exam. The board exam consists of multiple choice questions.

- Please write your answer so that there is only one correct answer among 5 options.
- Give an explanation
- Rate your confidence in your answer according to Likert performance with the following scales: 1 = do not trust [indicates that he/she does not know]; 2 = little confidence [i.e., maybe]; 3 = some interference; 4 = confidence [i.e., likely]; 5 = high confidence [stating the answer and explanation without doubt])
- Grading the question's difficulty level according to Likert performance with the following grades: 1=Very easy question, 2=Easy question, 3=Medium

question, 4=Difficult question, 5=Very difficult question.

Make sure you have these 4 in your output.

Answer each question with this format:

Answer:

Explanation:

Confidence level

Difficulty level:"

Statistical analyses and ethical considerations

The baseline question characteristics, and the accuracy of the LLMs were expressed with frequencies and percentages. The comparison of the accuracy of the individual LLMs and the comparison of the accuracy of the LLMs in original and translated versions of the questions were conducted with Chi-square tests. The comparison of median Likert scores for case questions vs general knowledge was conducted with Mann-Whitney U test. All statistical analyses were conducted with SPSS, version 25.0 (IBM Inc., Armonk, NY, USA), and a type 1 error level of 5% ($p<0.05$) was considered as the threshold limit for statistical significance.

Due to the use of a previously available question bank and no involvement of human subjects, the study is exempt from ethical approval. The study was conducted and reported according to STrengthening the Reporting of OBservational studies in Epidemiology (STROBE) guidelines [19].

Results

A total of 290 questions were evaluated via three LLMs. Breast (n=31), lung (n=24), and colorectal cancers (n=29) were the most frequently assessed tumor types in the question bank. The distribution of case questions and general knowledge-based questions were even (49.7% vs. 50.3%). The treatment was the most frequently evaluated area (49.3%). The topic distribution for the questions is summarized in Table 1.

The overall accuracy of LLMs was 59.3%, 42.8%, 36.2% for ChatGPT4, ChatGPT3.5, and Google Bard, respectively. The accuracy of ChatGPT4 was significantly higher than that of ChatGPT3.5 ($p<0.001$) and Google Bard ($p<0.001$), while the accuracy of ChatGPT3.5 was higher than that of Google Bard ($p<0.001$). Only the ChatGPT4 was proficient in all three examination years (2021-2023). The three LLMs had moderate correlation in their accuracy for individual responses ($r=0.319$, $p<0.001$ for ChatGPT 3.5 vs. ChatGPT 4, $r=0.263$, $p<0.001$ for ChatGPT 3.5 vs. Google Bard and $r=0.288$, $p<0.001$ for ChatGPT 4 vs. Google Bard). The accuracy of the LLMs was higher for the translated questions compared to original language in all three LLMs (62.8 vs. 59.3%, $p<0.001$ for ChatGPT 4, 48.3 vs. 42.8%, $p<0.001$ for ChatGPT3.5and 43.1 vs. 36.2%, $p<0.001$ for Google Bard). While the accuracy was improved with the translated questions, the LLMS were not able to replicate the correct answers given in the original language for over 10% of the questions (Table 2). The LLMs were more accurate in general knowledge than case questions (Table 3). The ChatGPT 3.5 ($p=0.301$) and Google Bard ($p=0.378$) had similar accuracy across variable knowledge domains (treatment, diagnosis,

Table 1. General characteristics of the question bank

Feature	n (%)
Question Type	
General Information	146 (50.3)
Case Question	144 (49.7)
Question Category	
Diagnosis	39 (13.4)
Treatment	143 (49.3)
Prognosis	32 (11)
Toxicity	22 (7.6)
General Information	54 (18.6)
Question Topic	
Basic Science	55 (19)
Breast	31 (10.7)
Lung	24 (8.3)
GI	51 (17.6)
GU	28 (9.7)
GYN	22 (7.6)
HNC	15 (5.2)
Sarcoma	19 (6.6)
Hematology	16 (5.5)
Other	29 (10)

Table 2. Comparison of individual LLMs performance

		Translated Questions (English)			
		ChatGPT 3.5			
Original Questions (Turkish)		**Wrong**	**Correct**	**Total**	**p value**
ChatGPT 3.5	Wrong n, (%)	122 (42.10)	44 (15.20)	166 (57.20)	<0.001
	Correct n, (%)	28 (9.70)	96 (33.10)	124 (42.80)	
	Total	150 (51.70)	140 (48.30)	290 (100)	
		ChatGPT 4			
		Wrong	**Correct**	**Total**	**p value**
ChatGPT 4	Wrong, n (%)	78 (26.90)	40 (13.80)	118 (40.70)	<0.001
	Correct, n (%)	30 (10.30)	142 (49.00)	172 (59.30)	
	Total	108 (37.20)	182 (62.80)	290 (100)	
		Google Bard			
		Wrong	**Correct**	**Total**	**p value**
Google Bard	Wrong, n (%)	122 (42.10)	63 (21.70)	185 (63.8)	<0.001
	Correct, n (%)	43 (14.80)	62 (21.40)	105 (36.2)	
	Total	165 (56.90)	125 (43.10)	290 (100)	

prognosis, toxicity, general information), while the performance of ChatGPT 4 varied across knowledge domains (Table 3). The accuracy of the LLMs across tumor types demonstrated similar accuracy across most tumor types.

The median Likert score was 4 for ChatGPT4, ChatGPT3.5, and Google Bard (Figure 1). The LLMs' answers' certainty was higher in general knowledge than in case questions for Google Bard (p=0.001), while the certainty for general knowledge questions and case questions were similar in ChatGPT 3.5 (p=0.135) and ChatGPT 4 (p=0.111). The median difficulty was 3/5 for all three LLMs (Figure 2). The question difficulty was regarded as higher case questions than general knowledge questions for ChatGPT 3.5 (p<0.001), ChatGPT 4 (p<0.001) and Google Bard (p<0.001).

Discussion

In the present study, we observed that LLMs had moderate success in medical oncology proficiency. The performance of the individual LLMs significantly varied, with ChatGPT4 outperforming the two other LLMs. The LLMs underperformed in clinical cases, mirroring complex clinical scenarios in daily practice. Additionally, the performance of LLMs was lower in the native language compared to questions translated to English. To best our knowledge, our study is the first study evaluating the proficiency of LLMs in oncology proficiency.

The use of LLMs as an assistant to clinical practice garnered a lot interest in the last year, especially after the eye opening performance of the ChatGPT 3 in USMLE examination [20]. In the pivotal study which was broadcasted via even television news, the ChatGPT achieved around 60% accuracy and an acceptable reasoning for the responses (20). In a later work, Brin et al. compared the performance of ChatGPT 4 and ChatGPT 3.5. The ChatGPT4 outperformed ChatGPT 3.5 (correct response rate 90 vs. 62.5%) [21]. Additionally, the ChatGPT4 was consistent with repeated evaluations, while the ChatGPT 3.5 revised its' responses in 82.5% of the cases [21], further supporting the use of ChatGPT 4 in medical knowledge. However, later studies challenged these findings.

Regarding the accuracy of LLMs in cancer care, similar unequivocal results exist. The ChatGPT 4 was accurate and comprehensive, with over 85% of questions related to head and neck cancer knowledge [22]. The questions in the study were generated from the frequently asked questions of professional societies, support groups, and

Table 3. Accuracy of individual LLMs across question types

		ChatGPT 3.5		ChatGPT 4		Google Bard	
		Wrong	Correct	Wrong	Correct	Wrong	Correct
		Count (%)	Count (%)	Count (%)	Count (%)	Count (%)	Count (%)
Question Type	General Information	70 (47.9)	76 (52.1)	41 (28.1)	105 (71.9)	85 (58.2)	61 (41.8)
	Case Question	96 (66.7)	48 (33.3)	77 (53.5)	67 (46.5)	100 (69.4)	44 (30.6)
Knowledge Domain	Diagnosis	19 (48.7)	20 (51.3)	6 (15.4)	33 (84.6)	23 (59)	16 (41)
	Treatment	87 (60.8)	56 (39.2)	74 (51.7)	69 (48.3)	98 (68.5)	45 (31.5)
	Prognosis	19 (59.4)	13 (40.6)	13 (40.6)	19 (59.4)	21 (65.6)	11 (34.4)
	Toxicity	15 (68.2)	7 (31.8)	11 (50)	11 (50)	14 (63.6)	8 (36.4)
	General Information	26 (48.1)	28 (51.9)	14 (25.9)	40 (74.1)	29 (53.7)	25 (46.3)
Question Topic	Breast	22 (71)	9 (29)	12 (38.7)	19 (61.3)	17 (54.8)	14 (45.2)
	Lung	13 (54.2)	11 (45.8)	10 (41.7)	14 (58.3)	16 (66.7)	8 (33.3)
	GI	27 (52.9)	24 (47.1)	24 (47.1)	27 (52.9)	38 (74.5)	13 (25.5)
	GU	18 (64.3)	10 (35.7)	17 (60.7)	11 (39.3)	19 (67.9)	9 (32.1)
	HNC	10 (66.7)	5 (33.3)	8 (53.3)	7 (46.7)	11 (73.3)	4 (26.7)
	GYN	14 (60)	8 (36.4)	6 (27.3)	16 (72.7)	12 (54.5)	10 (45.5)
	Sarcoma	5 (26.3)	14 (73.7)	7 (36.8)	12 (63.2)	10 (52.6)	9 (47.4)
	Basic Science	30 (54.5)	25 (45.5)	18 (32.7)	37 (67.3)	35 (63.6)	20 (36.4)
	Hematology	10 (62.5)	6 (37.5)	6 (37.5)	10 (62.5)	10 (62.5)	6 (37.5)
	Other	17 (58.6)	12 (41.4)	10 (34.5)	19 (65.5)	17 (58.6)	12 (41.4)

social media, reflecting the potential of the LLMs in population-level healthcare education [22]. In contrast, ChatGPT provided accurate and comprehensive responses to only 53.1% of the questions regarding cervical cancer care in another study. [23]. It should be noted that this study used ChatGPT 3.5 [23], which underperformed in our analysis compared to ChatGPT 4. Whether the newer versions of the LLMs could improve healthcare information should be investigated.

There are several caveats regarding the use of LLMs in cancer care. First, the LLMs gather information from several sources with various publishing dates. In the earlier works with the ChatGPT in healthcare, the training time limitation up to September 2021 was an important problem. While this issue was resolved in September 2023, some of the recommendations by LLMs are still outdated. Although the LLMs gather information from several sources, the reasoning for selecting a particular reference is not fully delineated by artificial intelligence, as previously noted in IBM Watson studies [24]. This issue is particularly problematic in complex scenarios with more than one therapeutic option. Additionally, the LLMs could not understand and solve the complex clinical scenarios that are very common in clinical practice [25]. Moreover, LLMs may generate errors in treatment sequencing and timing in such cases and may assign high confidence to incorrect or potentially unsafe recommendations. Furthermore, in line with the general principles of the LLMs, the LLMs used very confident language even in the obviously wrong simple questions [13,26]. The reasoning for the problems requiring the best option. Lastly, less is known about the comparative efficacy of different LLMs. In several studies, the performance of newer versions of the LLMs was better compared to previous versions [11,13], and the ChatGPT outperformed Google Bard in a very recent study on radiology exam questions [17]. Further studies should separately evaluate the performance of different LLMs to delineate the best model for individual scenarios.

In conclusion, the LLMs had below acceptable performance in a national oncology board exam with only the ChatGPT 4.0 had proficiency in an oncology

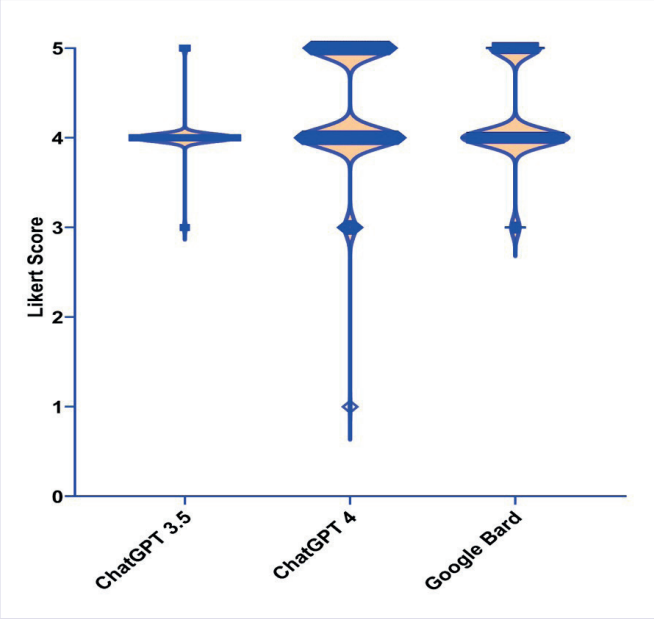


Figure 1. Confidence level rated on a 5-point Likert Scale (1 = very unconfident, 5 = very confident)

board examination. The LLMs' success in questions mirroring clinical practice was lower. Further research is needed to improve the proficiency of LLMs in cancer care-related information.

Author contributions

Conception and design: T.K.S., M.D., N.K., D.C.G.; Data acquisition: T.K.S., D.C.G.; Data analysis: T.K.S., D.C.G.; Data interpretation: T.K.S., D.C.G.; Drafting of the manuscript: T.K.S., M.D., N.K., D.C.G.; Critical revision of the manuscript: T.K.S., D.C.G.. All authors reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of this study.

Ethical approval

Due to the use of a previously available question bank and no involvement of human subjects, the study is exempt from ethical approval.

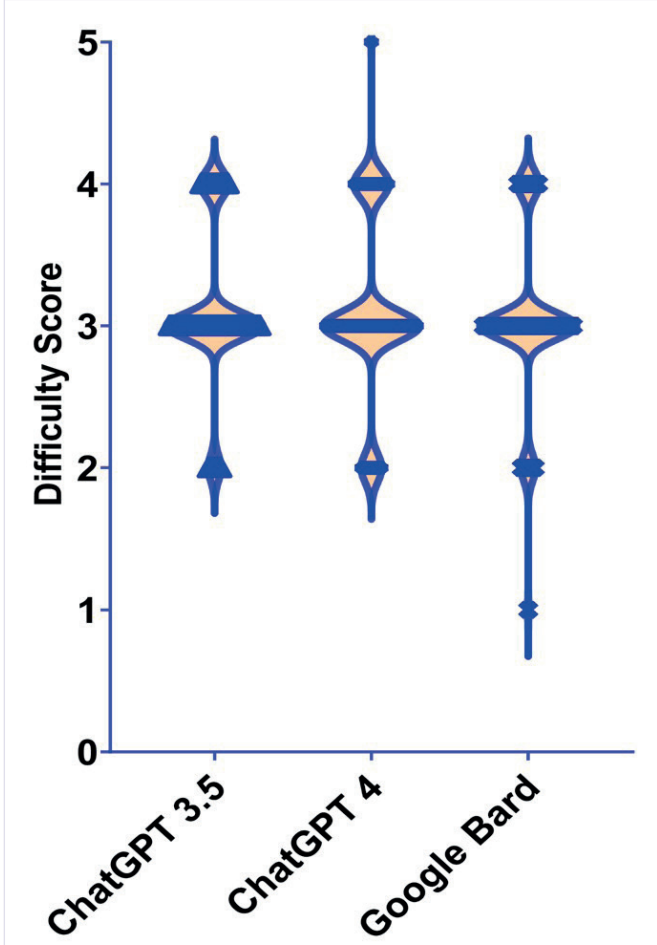


Figure 2. Difficulty level rated on a 5-point Likert Scale (1 = very easy, 5 = very difficult)

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The authors declare that this study received no funding.

Generative AI statement

The authors declare that no generative AI or AI-assisted technologies were used in the writing or preparation of this study.

References

[1] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25(1):44-56. **[Crossref]**

[2] Bhinder B, Gilvary C, Madhukar NS, Elemento O. Artificial intelligence in cancer research and precision medicine. Cancer Discov 2021;11(4):900-15. **[Crossref]**

[3] Elemento O, Leslie C, Lundin J, Tourassi G. Artificial intelligence in cancer research, diagnosis and therapy. Nat Rev Cancer 2021;21(12):747-52. **[Crossref]**

[4] Malin JL. Envisioning Watson as a rapid-learning system for oncology. J Oncol Pract 2013;9(3):155-7. **[Crossref]**

[5] Benary M, Wang XD, Schmidt M, et al. Leveraging Large Language Models for decision support in Personalized Oncology. JAMA Netw Open 2023;6(11):e2343689. **[Crossref]**

[6] Clusmann J, Kolbinger FR, Muti HS, et al. The future landscape of large language models in medicine. Commun Med (Lond) 2023;3(1):141. **[Crossref]**

[7] Khan RA, Jawaid M, Khan AR, Sajjad M. ChatGPT - Reshaping medical education and clinical management. Pak J Med Sci 2023;39(2):605-7. **[Crossref]**

[8] Javaid M, Haleem A, Singh RP. ChatGPT for healthcare services: an emerging stage for an innovative perspective. BenchCouncil Transactions on Benchmarks, Standards and Evaluations 2023;3(1):100105. **[Crossref]**

[9] Gilson A, Safranek CW, Huang T, et al. How does ChatGPT perform on the United States Medical Licensing Examination (USMLE)? The implications of Large Language Models for medical education and knowledge assessment. JMIR Med Educ 2023;9:e45312. **[Crossref]**

[10] Bhayana R, Krishna S, Bleakney RR. Performance of ChatGPT on a Radiology Board-style Examination: insights into current strengths and limitations. Radiology 2023;307(5):e230582. **[Crossref]**

[11] Antaki F, Touma S, Milad D, El-Khoury J, Duval R. Evaluating the performance of ChatGPT in Ophthalmology: an analysis of its successes and shortcomings. Ophthalmol Sci 2023;3(4):100324. **[Crossref]**

[12] Chen TC, Kaminski E, Koduri L, et al. Chat GPT as a Neuro-Score Calculator: analysis of a Large Language Model's performance on various neurological exam grading scales. World Neurosurg 2023;179:e342-e347. **[Crossref]**

[13] Schubert MC, Wick W, Venkataramani V. Performance of Large Language Models on a Neurology Board-Style Examination. JAMA Netw Open 2023;6(12):e2346721. **[Crossref]**

[14] Gunesli I, Aksun S, Fathelbab J, Yildiz BO. Comparative evaluation of ChatGPT-4, ChatGPT-3.5 and Google Gemini on PCOS assessment and management based on recommendations from the 2023 guideline. Endocrine 2025;88(1):315-22. **[Crossref]**

[15] Erul E, Aktekin Y, Danışman FB, et al. Perceptions, attitudes, and concerns on Artificial Intelligence applications in patients with cancer. Cancer Control 2025;32:10732748251343245. **[Crossref]**

[16] Skalidis I, Cagnina A, Luangphiphat W, et al. ChatGPT takes on the European Exam in Core Cardiology: an artificial intelligence success story? Eur Heart J Digit Health 2023;4(3):279-81. **[Crossref]**

[17] Patil NS, Huang RS, van der Pol CB, Larocque N. Comparative performance of ChatGPT and Bard in a text-based radiology knowledge assessment. Can Assoc Radiol J 2024;75(2):344-50. **[Crossref]**

[18] Ali R, Tang OY, Connolly ID, et al. Performance of ChatGPT and GPT-4 on neurosurgery written board examinations. Neurosurgery 2023;93(6):1353-65. **[Crossref]**

[19] von Elm E, Altman DG, Egger M, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. BMJ 2007;335(7624):806-8. **[Crossref]**

[20] Kung TH, Cheatham M, Medenilla A, et al. Performance of ChatGPT on USMLE: potential for AI-assisted medical education using large language models. PLOS Digit Health 2023;2(2):e0000198. **[Crossref]**

[21] Brin D, Sorin V, Vaid A, et al. Comparing ChatGPT and GPT-4 performance in USMLE soft skill assessments. Sci Rep 2023;13(1):16492. **[Crossref]**

[22] Kuşcu O, Pamuk AE, Sütay Süslü N, Hosal S. Is ChatGPT accurate and reliable in answering questions regarding head and neck cancer? Front Oncol 2023;13:1256459. **[Crossref]**

[23] Hermann CE, Patel JM, Boyd L, Growdon WB, Aviki E, Stasenko M. Let's chat about cervical cancer: assessing the accuracy of ChatGPT responses to cervical cancer questions. Gynecol Oncol 2023;179:164-8. **[Crossref]**

[24] Tupasela A, Di Nucci E. Concordance as evidence in the Watson for Oncology decision-support system. AI & SOCIETY 2020;35(4):811-8. **[Crossref]**

[25] Karabacak M, Margetis K. Embracing large language models for medical applications: opportunities and challenges. cureus. 2023;15(5):e39305. **[Crossref]**

[26] Rohrbach A, Hendricks LA, Burns K, Darrell T, Saenko K. Object hallucination in image captioning. Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, Association for Computational Linguistics; 2018, 4035-45. **[Crossref]**

acta medica

RESEARCH ARTICLE

Diagnostic challenges in suspected mitochondrial disease: Clinical, metabolic, and genetic findings

Doina Secu[1], Daniela Blanita[2], Natalia Usurelu[2], Victoria Sacara[1]

[1]Laboratory of Human Molecular Genetics, Institute of Mother and Child, Chisinau, Republic of Moldova
[2]Laboratory of Prevention of Hereditary Pathologies, Institute of Mother and Child, Chisinau, Republic of Moldova

Abstract

Objectives: This study delineates the diagnostic architecture of patients referred with suspected mitochondrial disease through the integrated analysis of clinical, biochemical, instrumental, and genetic data. By comparing patients with mitochondrial involvement, alternative genetic disorders, and unresolved cases, we aim to define phenotypic and molecular patterns associated with diagnostic stratification and characterize the diverse spectrum of genetic and non-genetic conditions that converge phenotypically on mitochondrial disease.

Materials and Methods: A total of 240 patients with clinical suspicion of mitochondrial disease were consecutively enrolled and assessed using a modified Nijmegen Mitochondrial Disease Score. All participants underwent comprehensive clinical, metabolic, instrumental, and neuroimaging evaluations, complemented by systematic molecular analyses, starting with common mitochondrial variant screening and progressing to more extensive targeted investigations guided by clinical and biochemical findings. Comparative analyses across the three groups employed nonparametric and categorical statistical tests.

Results: Relevant molecular findings were identified in 81 patients (33.7%), encompassing 37 (15.4%) with mitochondrial involvement and 44 (18.3%) with alternative genetic disorders, while 159 individuals (66.3%) remained unresolved. The mitochondrial group exhibited significantly higher rates of neuromuscular, brainstem, ophthalmic, and cardiac involvement, along with developmental regression, whereas seizures were a shared hallmark of both mitochondrial and non-mitochondrial subgroups. Biochemically, elevated serum lactate and plasma alanine were the most discriminative markers for the mitochondrial group, with a significantly higher prevalence of abnormal acylcarnitine profiles and organic aciduria in both the mitochondrial and alternative genetic subgroups. The undiagnosed cohort demonstrated phenotypic convergence with confirmed cases but lacked definitive molecular correlates despite extensive evaluation.

Conclusions: These findings underscore the intrinsic complexity and phenotypic heterogeneity of suspected mitochondrial disease, affirming the critical role of integrated clinical, metabolic, instrumental, and molecular assessment while simultaneously highlighting the limitations of current diagnostic paradigms and the imperative for expanded genomic and functional strategies to enhance resolution in unresolved cases.

Keywords: mitochondrial disease, mtDNA, nijmegen mitochondrial disease score, multisystem involvement, metabolic profilin

Corresponding author: Doina Secu • **Email:** secudoina95@gmail.com

Received: January 16, 2026 **Accepted:** May 3, 2026 **Published online:** June 28, 2026

Introduction

Mitochondrial diseases are a diverse group of inherited disorders caused by defects in mitochondrial oxidative phosphorylation, resulting in impaired cellular energy production. These conditions arise from pathogenic variants in either mitochondrial DNA (mtDNA) or nuclear genes encoding proteins essential for mitochondrial structure and function. The dual genetic origin of these disorders, together with variable heteroplasmy, the coexistence of mutant and wild-type mtDNA within the same cell, and complex genotype-phenotype relationships, contributes to marked clinical variability and a wide spectrum of disease severity, with onset ranging from childhood to adulthood [1].

Clinically, organs with high energy requirements, including the central nervous system, skeletal muscle, heart, and endocrine system, are most commonly affected. However, mitochondrial diseases lack pathognomonic clinical features and frequently share overlapping manifestations with other genetic, metabolic, and neuromuscular disorders. This substantial phenotypic overlap represents a major challenge in the diagnostic evaluation of patients with suspected mitochondrial disease and often leads to delayed, incomplete, or incorrect diagnoses [2].

The estimated prevalence of mitochondrial diseases is approximately 1 in 5,000 individuals, placing them among the most common inherited metabolic disorders. When patients with suspected but unconfirmed disease are considered, the true prevalence is likely higher, further underscoring the clinical and diagnostic burden associated with these conditions. Despite significant advances in molecular genetic techniques, a definitive diagnosis cannot be established in a considerable proportion of patients [3].

The diagnostic workup of suspected mitochondrial disease requires an integrated, multidisciplinary approach, combining clinical evaluation with biochemical, instrumental, neuroimaging, and genetic investigations. Although clinical scoring systems, such as the Nijmegen Mitochondrial Disease Score, are useful for diagnostic stratification, they lack disease specificity and may capture a broad range of non-mitochondrial conditions with overlapping phenotypes [4]. Consequently, even after extensive diagnostic assessment, many patients remain without a conclusive diagnosis.

The aim of the present study was to provide a comprehensive comparative characterization of patients with suspected mitochondrial disease, stratified by molecular outcome into: mitochondrial involvement, alternative genetic disorders, and unresolved cases despite extensive clinical and molecular evaluation. Through a multidimensional analysis of clinical, biochemical, instrumental, neuroimaging, and genetic features, we aimed to identify distinguishing profiles across these groups, to highlight the limitations of current diagnostic strategies, and to provide a framework for improving future diagnostic pathways in mitochondrial medicine.

Materials and methods

Enrollment strategy and cohort characterization

A cohort of 240 patients with clinical suspicion of mitochondrial disease was consecutively enrolled at the Institute of Mother and Child, a tertiary care hospital in the Republic of Moldova, between March 2021 and October 2024. Patients were primarily referred to our unit from various specialized departments, including neurology, cardiology, ophthalmology, and other clinical units, due to complex, multisystemic presentations that remained undiagnosed after routine investigations.

Upon referral, each patient underwent a standardized clinical assessment by our multidisciplinary team to objectively quantify the level of suspicion using a modified Nijmegen Mitochondrial Disease Score (NMDS) [5] (Table 1).

The NMDS is a validated domain-based tool that integrates clinical features, metabolic and biochemical data, and neuroimaging findings to estimate the likelihood of mitochondrial involvement. In its original form, the score also incorporates muscle biopsy and enzymatic assessments, which were not available for all patients in this cohort. To adapt to these constraints, the biopsy and enzymology domain was omitted, and the score was calculated solely based on clinical, biochemical, and neuroimaging criteria, maintaining the original scoring structure within these domains. Scores are interpreted as follows: 1 indicates mitochondrial disorder unlikely, 2-4 suggests a possible mitochondrial disorder, 5-7 corresponds to a probable mitochondrial

Table 1. Modified nijmegen mitochondrial disease score

I. Clinical signs and symptoms (max. 4 points)			II. Metabolic/imaging studies (max. 4 points)
Muscular presentation (max. 2 points)	**CNS presentation (max. 2 points)**	**Multisystem disease (max. 3 points)**	
Ophthalmoplegia †	Developmental delay	Hematology	Elevated lactate †
Facies myopathica	Loss of skills	Gastrointestinal tract	Elevated lactate/pyruvate ratio
Exercise intolerance	Stroke-like episode	Endocrine/growth Heart	Elevated alanine †
Muscle weakness	Migraine	Kidney	Elevated CSF lactate †
Rhabdomyolysis	Seizures	Vision	Elevated CSF protein
Abnormal electromyography	Myoclonus	Hearing	Urinary tricarbon acid excretion †
	Cortical blindness	Neuropathy	Elevated CSF alanine †
	Pyramidal signs	Recurrent/familial	Ethylmalonic aciduria
	Extrapyramidal signs		Stroke-like picture/MRI
	Brainstem involvement		Leigh syndrome/MRI †
			Elevated lactate/MRS

Legend: † – indicates that this specific symptom scores 2 points; CSF – cerebrospinal fluid; MRI – magnetic resonance imaging; MRS – magnetic resonance spectroscopy.

disorder, and 8 reflects a definite mitochondrial disorder. An inclusion threshold of ≥3 points was retained to capture individuals with possible mitochondrial disease, ensuring that patients with early or partially expressed phenotypes were not excluded.

The study was approved by the Research Ethics Committee of the State University of Medicine and Pharmacy Nicolae Testemițanu and was conducted in accordance with the Declaration of Helsinki. Written informed consent was obtained from all participants or their legal guardians prior to inclusion in the study.

Patients were subsequently stratified into three categories based on diagnostic outcomes: individuals with confirmed mitochondrial involvement (n = 37), those with alternative genetic disorders (n = 44), and patients who remained without a definitive diagnosis (n = 159). Within this framework, the mitochondrial involvement group was defined by the identification of pathogenic or likely pathogenic variants, alongside specific variants of uncertain significance (VUS), including homoplasmic mtDNA mutations, that directly implicate the oxidative phosphorylation machinery. While these genetic findings provided a robust molecular explanation for most cases, this classification also encompassed patients where the detected variants offered a substantial causal contribution to the clinical presentation, serving as a foundational diagnostic anchor even in cases where they provided only a partial explanation for an exceptionally complex phenotype. Conversely, the alternative genetic disorders group was operationalized as an etiologically distinct collective originating from non-mitochondrial pathways, such as ion channelopathies or various syndromic conditions, which exhibit a marked phenotypic convergence with mitochondrial disease by replicating its multisystemic and complex clinical trajectory. Finally, the unresolved group comprised individuals who met the inclusion criteria but in whom no causative genetic variants were identified, representing the inherent diagnostic limitations in current genomic screening. In the following sections, we detail our institutional experience across these three cohorts, providing a comprehensive appraisal of the clinical, biochemical, and molecular patterns observed to improve the differential diagnosis between confirmed or suspected mitochondrial involvement and its genetic mimics.

Clinical, familial, and paraclinical evaluation

All participants underwent a thorough clinical assessment, including detailed medical history, family history, prenatal and perinatal data, and age at symptom onset. Systematic physical and neurological examinations were conducted to evaluate multisystem

involvement. Routine laboratory testing encompassed hematological indices, renal and hepatic function, serum electrolytes, lactate, creatine kinase, and lactate dehydrogenase.

Extended metabolic investigations, conducted for a subset of the cohort, encompassed plasma amino acid profiling via high-performance liquid chromatography, acylcarnitine quantification by liquid chromatography-tandem mass spectrometry, and urinary organic acid analysis using nuclear magnetic resonance spectroscopy. Pathological status was defined by persistent deviations from laboratory-established reference intervals, encompassing disproportionate accumulations of short-, medium-, and long-chain acylcarnitine species, aberrant diagnostic ratios indicative of impaired intermediary metabolism, or the abnormal excretion of organic acid biomarkers, including dicarboxylic, methylmalonic, and glutaric acids, alongside elevated Krebs cycle intermediates.

Instrumental and imaging studies were performed based on clinical indications, comprising electroencephalography, electromyography, electrocardiography, audiometry, and cerebral imaging (MRI or CT) to detect structural or functional abnormalities.

Genetic analysis

All 240 patients underwent initial molecular prescreening for seven recurrent pathogenic mtDNA point mutations, specifically m.3243A>G, m.8344A>G, m.8993T>G/C, m.13513G>A, m.3460G>A, m.11778G>A, and m.14484T>C, using quantitative polymerase chain reaction with high-resolution melting analysis (qPCR-HRM). To ensure diagnostic precision and calibrate melting profiles, synthetic oligonucleotide controls representing both wild-type and mutant genotypes were systematically employed. Heteroplasmy levels for variants identified via qPCR-HRM were further evaluated using PCR-restriction fragment length polymorphism, with semi-quantitative assessment performed through densitometric analysis of electrophoresis gels using ImageJ software.

Patients with a modified NMDS ≥6, as well as those exhibiting inconclusive results in the prescreening, were advanced to targeted Sanger sequencing of the mitochondrial genome. This threshold of ≥6 was strategically established to prioritize molecular resources for a subgroup with the highest pre-test probability of mitochondrial disease. Within the NMDS framework, a score exceeding 6 ensures that clinical signs (which contribute a maximum of 4 points in Category I) are supported by objective findings from Category II, such as hyperlactatemia, hyperalaninemia, or specific mitochondrial signatures on brain MRI. The analysis covered all 13 protein-coding genes and included flanking tRNA and rRNA genes adjacent to the coding regions, ensuring comprehensive assessment of loci relevant to mitochondrial disease. Detected variants were interpreted according to the American College of Medical Genetics and Genomics (ACMG) guidelines [6], ensuring a standardized and clinically meaningful classification framework.

A subset of patients underwent additional molecular analyses of the nuclear genome using targeted Sanger sequencing and next-generation sequencing (NGS) technologies, including whole-exome sequencing (WES), whole-genome sequencing (WGS) and targeted multigene panels. These genomic investigations were complemented in some instances by Multiplex Ligation-dependent Probe Amplification (MLPA) or array comparative genomic hybridization (aCGH). All procedures were conducted in ISO 15189-accredited laboratories, both domestically and internationally, ensuring methodological rigor, reproducibility, and adherence to recognized quality standards. These comprehensive analyses facilitated the identification of nuclear gene variants implicated in mitochondrial dysfunction, as well as the detection of variants unrelated to mitochondrial disease. Detected variants were systematically interpreted and classified according to the ACMG guidelines, considering pathogenicity, allele frequency, predicted functional impact, and available clinical evidence, thereby providing a standardized and clinically meaningful framework for molecular diagnosis.

DNA extraction for all molecular investigations, including qPCR-HRM screening, Sanger sequencing, and NGS-based analyses, was performed exclusively from peripheral blood samples.

Statistical analysis

Statistical analyses were performed using SPSS version 28.0 (IBM Corp., USA). Continuous variables were expressed as mean ± standard deviation or median with interquartile range, whereas categorical variables were

summarized as counts and percentages. Comparisons across the three investigation groups were conducted using the Kruskal-Wallis test for continuous variables and the Chi-square test for categorical variables. A significance threshold of $p < 0.05$ was applied.

Results

The study population comprised 240 patients with suspected mitochondrial disease, stratified into three groups based on molecular outcomes: 37 (15.4%) with mitochondrial involvement, 44 (18.3%) with alternative genetic disorders, and 159 patients (66.3%) remained without a confirmed diagnosis.

The median age at study inclusion was 24 months (IQR: 11–68) for the mitochondrial involvement group, 24 months (IQR: 12–72) for alternative genetic disorders, and 30 months (IQR: 11–66) in the undiagnosed group, indicating substantial overlap in age distribution between the cohorts. Symptom onset occurred predominantly in early infancy, with a median age of 3 months (IQR: 1–15) in the mitochondrial group, 4 months (IQR: 1–10) in the alternative disorders group, and 2 months (IQR: 0–12) in undiagnosed patients.

No statistically significant differences were observed across the three investigation groups with respect to the presence of a positive family history of similar clinical manifestations. Additionally, the frequency of pregnancy and birth-related complications, including preterm delivery, intrauterine growth restriction, and other perinatal adversities, was comparable across groups.

Clinical manifestations

The clinical profile of the study cohort was highly heterogeneous, reflecting the multisystemic nature of suspected mitochondrial dysfunction. This diversity provided a structured framework for a comparative analysis of clinical prevalence and severity across the three investigation groups, with the distribution of these features among patients with mitochondrial involvement, alternative genetic disorders, and unresolved cases illustrated in Figure 1.

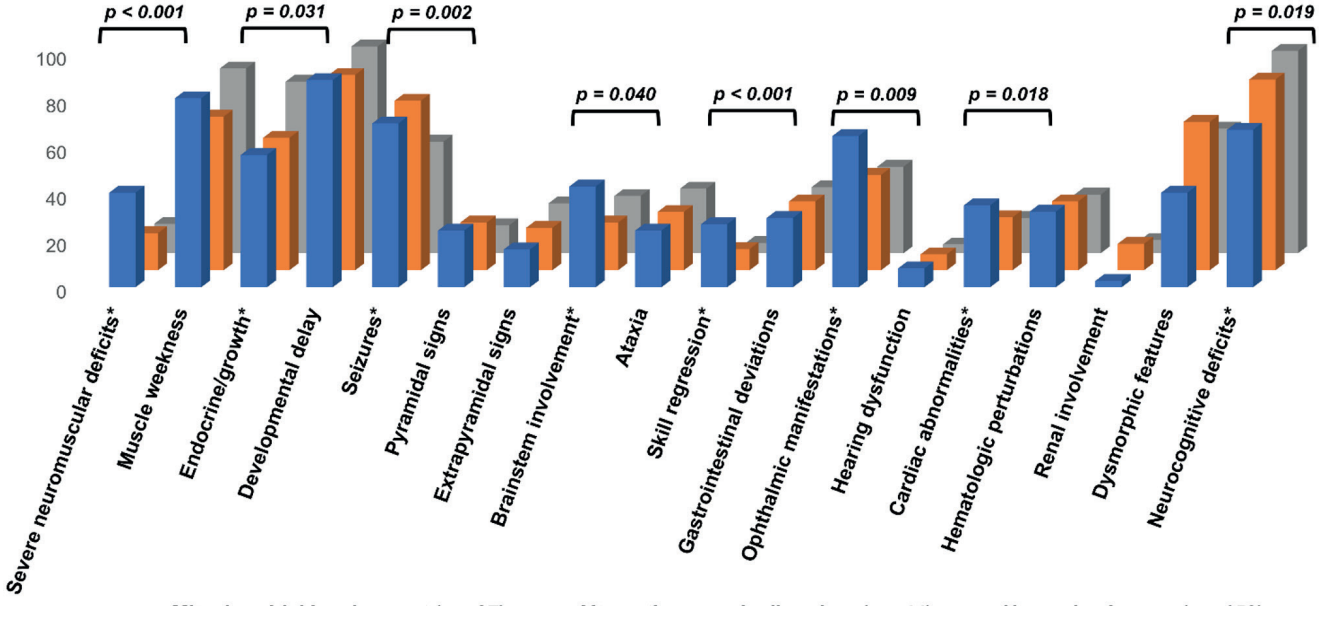


Figure 1. Prevalence of clinical features across patients with mitochondrial involvement, alternative genetic disorders, and unresolved cases

Legend: * – variables marked with an asterisk indicate statistically significant findings.

The mitochondrial involvement group exhibited significantly higher rates of severe neuromuscular deficits (40.5% vs. 15.9% and 12.6%; $p < 0.001$), brainstem involvement (43.2% vs. 20.5% and 24.5%; $p = 0.040$), ophthalmic manifestations (64.9% vs. 40.9% and 37.1%; $p = 0.009$), cardiac abnormalities (35.1% vs. 22.7% and 15.1%; $p = 0.018$), and skill regression (27.0% vs. 9.1% and 4.4%; $p < 0.001$) compared to the alternative genetic disorders and unresolved groups, respectively. Seizures were a shared predominant feature in both the mitochondrial involvement (70.3%) and alternative genetic disorders (72.7%) groups, occurring significantly more frequently than in the unresolved cohort (47.8%; $p = 0.002$).

Conversely, undiagnosed patients exhibited a higher prevalence of neurodevelopmental and behavioral impairments (86.8% vs. 67.6% and 81.8%; $p = 0.021$) endocrine or growth-related abnormalities (73.6% vs. 56.8% and 56.8%; $p = 0.011$), relative to the mitochondrial involvement and alternative genetic disorders groups, respectively.

Biochemical and instrumental investigations

Biochemical testing revealed that elevated serum lactate emerged as a discriminative marker between groups, being both significantly more prevalent as an abnormal finding in patients with mitochondrial involvement (34/37, 91.9%) compared to both the alternative genetic disorders (16/38, 42.1%) and unresolved (57/138, 41.3%) cohorts. Furthermore, a comparison of absolute concentrations using the Kruskal-Wallis test demonstrated that lactate levels were significantly higher in the mitochondrial group (Mean Rank: 159.6) than in the alternative (Mean Rank: 95.0) and unresolved (Mean Rank: 96.2) groups ($H = 32.66$, $p < 0.001$). Plasma alanine demonstrated a similar discriminative pattern, with significant differences observed in both prevalence ($p < 0.001$) and absolute concentrations across the three investigation groups ($H = 22.92$, $p < 0.001$). This significance was primarily driven by the mitochondrial involvement group, which presented elevated alanine levels in 60.0% (18/30) of patients, substantially exceeding the frequencies observed in the alternative genetic disorders (23.5%, 8/34) and unresolved (11.2%, 13/116) cohorts (Mean Ranks: 87.1, 59.0, and 46.5, respectively).

Additional metabolic screening identified significant disparities in secondary markers of metabolic dysfunction, revealing that pathological acylcarnitine profiles ($p = 0.027$) were notably more frequent in the alternative genetic disorders group (32.3%, 10/31) compared to the mitochondrial involvement (25.0%, 5/20) and unresolved (12.7%, 15/118) cohorts. Furthermore, while infrequent across all groups, organic aciduria ($p = 0.008$) was more common in the mitochondrial involvement (16.7%, 5/30) and alternative genetic disorders (13.5%, 5/37) groups than in the unresolved cohort (3.1%, 4/130).

Other routine laboratory parameters, including serum transaminases, creatine kinase, lactate dehydrogenase, and serum electrolytes, did not differ significantly between groups. The distribution of biochemical, instrumental, and neuroimaging abnormalities across the mitochondrial involvement, alternative genetic disorders, and unresolved groups is summarized in Figure 2.

Instrumental assessments revealed marked distinctions between cohorts. Electromyographic abnormalities reached their highest frequency in the alternative genetic disorders group (68.8%, 11/16), followed by the mitochondrial involvement group (50.0%, 9/18), with both cohorts showing significantly higher rates than the 31.3% (20/64) identified in the unresolved cohort ($p = 0.016$). This was paralleled by a higher frequency of electroencephalographic alterations in the alternative genetic (80.6%, 29/36) and mitochondrial (67.7%, 21/31) groups compared to unresolved cases (55.8%, 58/104; $p = 0.025$), alongside cardiographic deviations which were notably more prevalent in the mitochondrial involvement group (36.7%, 11/30) relative to the alternative genetic (18.9%, 7/37) and unresolved (13.4%, 17/127) cohorts ($p = 0.011$).

Neuroimaging abnormalities were significantly more prevalent in the mitochondrial involvement group (82.4%, 28/34) compared to the alternative genetic disorders (54.1%, 20/37) and unresolved (58.0%, 58/100) cohorts ($p = 0.022$). Specifically, the mitochondrial involvement group exhibited higher rates of cerebral or cerebellar atrophy (44.1% vs. 18.9% and 23.0%; $p = 0.026$) and basal ganglia involvement (17.6% vs. 2.7% and 2.0%; $p < 0.001$), underscoring a distinct neuroanatomical pattern of mitochondrial pathology.

The cumulative burden of these multi-systemic findings was further reflected in the clinical stratification provided by the NMDS. Within the mitochondrial involvement

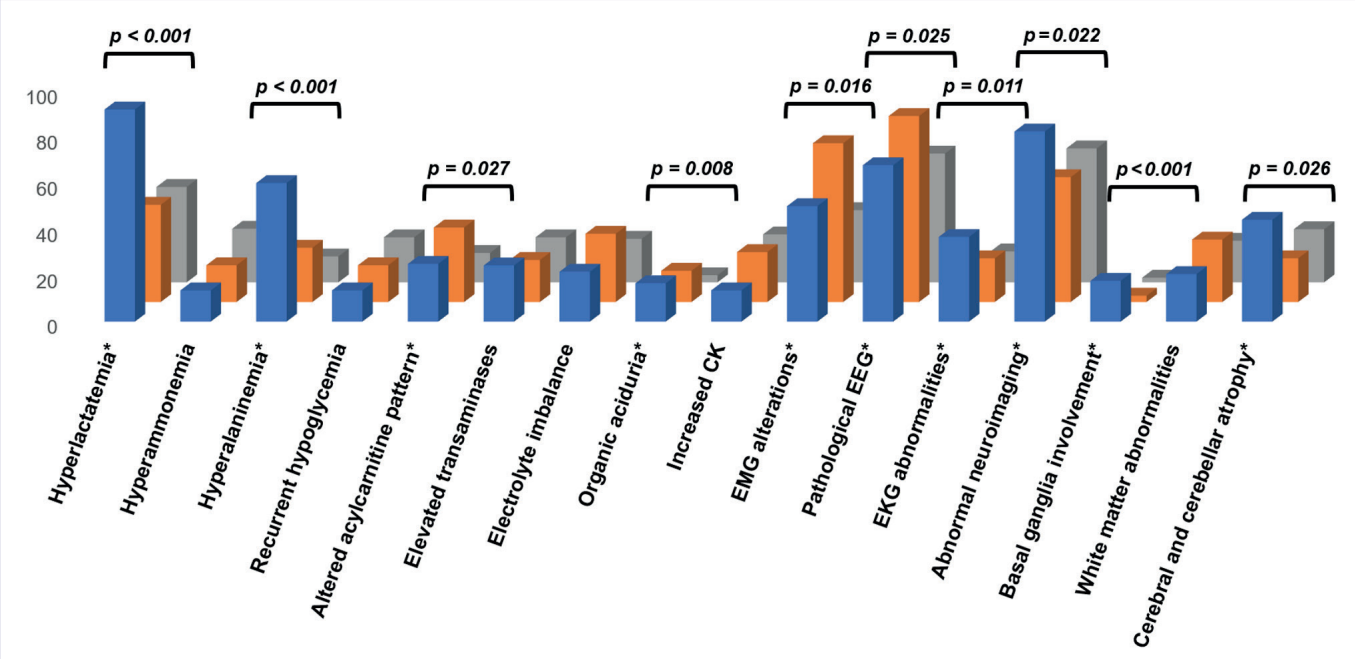


Figure 2. Prevalence of biochemical, instrumental, and neuroimaging abnormalities across patients with mitochondrial involvement, alternative genetic disorders, and unresolved cases

Legend: * – variables marked with an asterisk indicate statistically significant findings; CK – creatine kinase; EEG – electroencephalography; EKG – electrocardiography; EMG – electromyography.

group, clinical suspicion was most intense, with 10 patients reaching a definite mitochondrial disease classification and 26 a probable one, compared to only 1 possible case. In the alternative genetic disorders group, the distribution shifted toward milder clinical features, involving 3 definite, 17 probable, and 24 possible cases. In contrast, the unresolved cohort was heavily skewed toward lower clinical suspicion, being dominated by 98 possible and 59 probable qualifiers, with only 2 patients meeting the definite mitochondrial disease criteria. Ultimately, significantly higher NMDS scores were concentrated in the mitochondrial group ($p < 0.001$), confirming the scale's effectiveness in identifying individuals whose multi-organ clinical burden strongly aligns with a mitochondrial-related phenotype.

Genetic investigations

Genetic evaluation contributed decisively to case stratification within the study population. All patients underwent initial molecular screening using a qPCR-HRM-based approach targeting common pathogenic mtDNA variants. Through this first-line screening, eight patients were identified as carriers of established pathogenic mtDNA point mutations. Specifically, the m.3243A>G variant was detected in three patients, m.8993T>G in two patients, m.3460G>A in two patients, and m.11778G>A in one patient, all of whom were classified as molecularly confirmed mitochondrial cases.

Subsequently, a total of 82 patients with increased clinical suspicion, defined by an NMDS ≥6, underwent extended molecular analysis of the mitochondrial genome by targeted Sanger sequencing with capillary electrophoresis. This approach led to the identification of 24 additional patients harboring mtDNA variants: 11 individuals (45.8%) with variants classified as pathogenic or likely pathogenic and 13 patients (54.2%) with VUS. Although these VUS cases do not fulfill the criteria for formal molecular confirmation, they were included in the mitochondrial involvement group based on the presence of homoplasmic variants that closely correlated with the specific clinical phenotypes of the patients. Analysis of the functional distribution of these variants revealed a predominant impact on respiratory chain architecture, with mutations affecting Complex I in 25.0% of cases, followed by alterations involving Complex V in 12.5% and mitochondrial RNA genes in

16.7%. In contrast, variants affecting Complexes III and IV were comparatively infrequent, each accounting for 4.1% of cases, while a substantial proportion of patients (37.6%) harbored variants predicted to compromise multiple respiratory chain complexes, underscoring the extensive and functionally heterogeneous nature of mitochondrial genomic involvement within this clinically selected subgroup.

In selected patients presenting with sustained clinical features suggestive of mitochondrial involvement, extended nuclear genetic testing was undertaken to further elucidate the underlying molecular etiology. This diagnostic effort, utilizing targeted Sanger sequencing (n = 2) and WES (n = 3), yielded significant findings in five individuals. These investigations identified pathogenic or likely pathogenic variants in *POLG* (n = 2), *TWNK*, and *ETHE1*, alongside a VUS in the *DGUOK* gene. Notably, one patient evaluated via WES demonstrated dual molecular involvement, harboring a pathogenic mtDNA alteration, previously detected by qPCR-HRM, concomitant with a VUS in the *OPA1* gene.

Altogether, the integration of these findings resulted in a total of 37 patients with molecular evidence of mitochondrial involvement, encompassing both mitochondrial and nuclear genetic determinants. Detailed profiles of the identified genetic variants and corresponding patient phenotypes are summarized in Supplementary Table 1.

Among patients with non-mitochondrial conditions, genetic testing identified alternative molecular diagnoses in 44 cases, comprising 32 individuals with pathogenic or likely pathogenic variants and 12 with VUS. The diagnostic workflow utilized a multimodal approach primarily leveraging high-throughput sequencing (WGS, n = 15; WES, n = 14; targeted multigene panels, n = 6), supplemented by structural variant analysis (MLPA, n = 4; aCGH, n = 1), and targeted Sanger sequencing (n = 4). These findings were distributed across five principal diagnostic categories. Syndromic and sensory disorders represented the largest group (n = 15, 34.1%), with variants in genes associated with neurodevelopmental and multisystem syndromes, including *FOXG1*, *NSD1*, *PPP2R5D* and others. The inborn errors of metabolism group (n = 12, 27.3%) involved defects in amino acid, carbohydrate, lipid, and peroxisomal metabolism, exemplified by variants in *GLB1*, *PCCA*, *ALDOB*, and additional genes. Channelopathies (n = 8, 18.1%) were associated with variants in voltage-gated ion channel genes such as *SCN1A*, *SCN2A*, and *SCN10A*. Neuromuscular disorders (n = 4, 9.1%) included variants in *ANO5*, *DARS2*, and *TCAP*, affecting muscle integrity and protein translation. Finally, chromosomal microdeletion syndromes (n = 5, 11.4%), encompassed recurrent pathogenic regions including 4p16.3, 7q11.23, and 15q11–q13, among others, further illustrating the broad spectrum of genetic conditions that can phenotypically mimic suspected mitochondrial disease. The molecular characteristics of the identified variants and their associated clinical phenotypes are systematically presented in Supplementary Table 2.

Patients who remained without a definitive molecular diagnosis were subjected to the same initial qPCR-HRM screening; however, no pathogenic mtDNA variants fulfilling diagnostic criteria were detected, and further genetic evaluations, encompassing the targeted analyses employed in this study, did not yield conclusive molecular evidence, underscoring the limitations of current methodologies in resolving these complex cases.

Discussion

We evaluated a pediatric cohort with suspected mitochondrial disease by integrating clinical, biochemical, instrumental, and genetic data. By comparatively analyzing patients across the categories of confirmed mitochondrial involvement, alternative genetic disorders, and unresolved cases, our findings pinpoint specific phenotypic patterns that improve patient classification and reflect the diverse nature of mitochondrial pathology.

Despite broadly comparable demographic and perinatal characteristics, patients with mitochondrial involvement exhibited a markedly more severe and multisystemic phenotype. This subgroup demonstrated higher rates of severe neuromuscular dysfunction, brainstem involvement, ophthalmic manifestations, cardiovascular abnormalities, and developmental regression. These clinical features reflect the heightened susceptibility of high-energy-dependent tissues to mitochondrial impairment and confirm that extensive multisystemic involvement remains a key indicator of mitochondrial disease [7-8]. Seizures were a predominant feature in both the mitochondrial involvement and alternative

genetic disorders groups. Conversely, patients lacking molecular confirmation were predominantly distinguished by neurodevelopmental, behavioral, and growth-related abnormalities, which, while often prompting suspicion of mitochondrial dysfunction, are more frequently associated with alternative genetic etiologies exhibiting overlapping phenotypic features.

Biochemical profiling further distinguished these groups, with elevated serum lactate and plasma alanine emerging as the most discriminative metabolic markers for the mitochondrial involvement group, indicative of impaired oxidative phosphorylation and secondary perturbations in intermediary metabolism. Pathological acylcarnitine profiles and organic aciduria were more frequently observed in both the mitochondrial involvement and alternative genetic disorders subgroups. These findings highlight the importance of expanded biochemical screening for identifying both the metabolic effects of mitochondrial failure and other underlying systemic metabolic abnormalities.

The Nijmegen Mitochondrial Disease Score demonstrated strong utility in capturing multisystem involvement, with significantly higher values in the mitochondrial involvement group ($p < 0.001$). Given that relatively high scores were also noted in the alternative genetic group, in our setting, the NMDS effectively identified cases more likely to yield relevant molecular findings across the entire cohort.

Comprehensive genetic analyses, encompassing both targeted mitochondrial and selected nuclear testing, identified relevant molecular findings in 33.7% of patients. This approach effectively stratified patients according to molecular etiology and highlighted the critical role of integrated genomic evaluation in the diagnostic workflow of suspected mitochondrial disease. Within this context, the diagnostic distribution observed in the present study is concordant with that reported in previous investigations of large and medium-sized cohorts of patients referred with suspected mitochondrial disease, which consistently demonstrate a heterogeneous composition including confirmed mitochondrial disorders, alternative non-mitochondrial genetic conditions, and a substantial proportion of unresolved cases. In the largest series reported to date, Rouzier et al. [9] analyzed a cohort exceeding 2,000 patients and established a molecular diagnosis in approximately 20% of cases, identifying 322 patients with mitochondrial disease and 75 with non-mitochondrial genetic etiologies, while the majority of individuals remained without a definitive diagnosis. Similarly, van der Ven et al. [10] investigated 491 patients and achieved a diagnostic yield of 51% (n = 250), of whom 221 patients (45%) were diagnosed with non-mitochondrial disorders and 29 patients (6%) with mitochondrial disease, whereas 241 patients (49%) remained undiagnosed. Comparable results were reported by Kerr et al. [11], who evaluated 390 patients and established a molecular diagnosis in 184 cases (47.2%), including 115 patients (29.5%) with non-mitochondrial genetic disorders and 69 patients (17.7%) with mitochondrial disease, with 206 patients (52.8%) remaining without molecular resolution. At the family-based level, Schon et al. [12] analyzed 319 families and identified a genetic diagnosis in 104 families (32.6%), comprising 65 families (20.4%) with non-mitochondrial causes and 39 families (12.2%) with mitochondrial disease, while 215 families (67.4%) remained without a confirmed molecular diagnosis. In smaller cohorts, such as that reported by Grigalionienė et al. [13], diagnostic yields were similarly limited; among 83 patients, only 18 (21.7%) received a molecular diagnosis, including 11 patients (13.3%) with mitochondrial disease, six patients (7.2%) with non-mitochondrial disorders, and one patient (1.2%) with combined mitochondrial and other genetic pathology, leaving 65 patients (78.3%) undiagnosed. Collectively, these data place our findings within a robust and reproducible diagnostic paradigm characteristic of suspected mitochondrial disease.

While this study provides a comprehensive evaluation, several methodological limitations merit consideration when interpreting the findings. Although the cohort benefited from extensive clinical, biochemical, and instrumental characterization, invasive functional investigations, most notably muscle biopsy with histopathological, histochemical, and respiratory chain enzymatic analyses, could not be systematically performed, precluding direct tissue-level confirmation of mitochondrial dysfunction and potentially limiting diagnostic resolution in selected cases. Furthermore, genetic investigations were conducted exclusively using peripheral blood samples. While this non-invasive approach is standard for primary screening and provides adequate DNA yield, it may underrepresent tissue-specific heteroplasmy or fail to detect certain mitochondrial DNA alterations that are more reliably identified in post-mitotic tissues, such as muscle. Additionally, the use of Sanger sequencing for mtDNA analysis precluded the precise quantification of

heteroplasmy levels; however, the majority of variants we identified appeared homoplasmic, aligning with the expected molecular profile for these specific alterations.

Moreover, although the genetic diagnostic strategy was structured and clinically guided, it relied primarily on targeted assays, and the limited use of NGS constrained comprehensive interrogation of both nuclear and mitochondrial genomes. This methodological constraint is especially relevant considering that expanded NGS approaches, including whole exome and dual genome analyses, have been shown to substantially increase the diagnostic yield in suspected mitochondrial disorders [14-16]. In our setting, the restricted application of these advanced technologies reflects persistent regional challenges; as high-throughput genomic testing is not yet integrated into the national healthcare reimbursement system, its implementation remains largely dependent on the financial resources of the patients' families. Consequently, pathogenic variants beyond the scope of the applied methods may have remained undetected, contributing to the persistence of unresolved cases. Finally, the pediatric, referral-based nature of the cohort may constrain the generalizability of these findings to adult populations or unselected clinical settings. Nonetheless, these constraints mirror real-world diagnostic practice and highlight the continued need for integrated functional approaches and broader genomic strategies to improve diagnostic resolution in patients with suspected mitochondrial disease.

Conclusion

This study delivers a rigorous, multidimensional characterization of a pediatric cohort with suspected mitochondrial disease, integrating clinical, biochemical, instrumental, and genetic data to refine diagnostic stratification. Relevant molecular findings were identified in 33.7% of patients (n = 81), distributed between patients with mitochondrial involvement (n = 37) and alternative genetic disorders (n = 44), while 66.3% (n = 159) remained unresolved. Specifically, patients with mitochondrial involvement demonstrated pronounced multisystemic involvement, distinctive metabolic perturbations, and convergent neuroanatomical alterations, reflecting the complex pathophysiology and systemic impact of mitochondrial dysfunction. The results underscore the critical importance of integrated diagnostic frameworks that combine detailed clinical assessment with targeted and expanded genomic analyses. Beyond current strategies, there remains a clear necessity for advanced sequencing technologies and functional investigations to enhance diagnostic resolution and support precision management in complex clinical settings.

Author contribution

Conception and design: D.S., N.U., V.S.; Data acquisition: D.S., D.B., N.U., V.S.; Data analysis: D.S., D.B.; Data interpretation: D.S., D.B., V.S.; Drafting of the manuscript: D.S., D.B., V.S.; Critical revision of the manuscript: D.S., D.B., N.U., V.S. All authors reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of this study.

Ethical approval

This study was approved by the Research Ethics Committee of the State University of Medicine and Pharmacy "Nicolae Testemițanu" (Date: September 9, 2020, Decision/Protocol No: 03). Informed consent was obtained from all participants involved in this study.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

This study was supported by the institutional research project "Diagnosis and Monitoring of Genetic Diseases in the Prevention of Maternal and Child Health Disorders" (DiMoGEN, 140102) and by the project

"Genomic Medicine and Metabolomics Research in the Service of Prophylaxis of Genetic Diseases for Healthy Generations in the Republic of Moldova" (SCREENGEN, 20.80009.8007.22).

Generative AI statement

The authors declare that during the preparation of this study, the following AI-assisted technology was used: ChatGPT-5.5 on December 2025 - January 2026. Extent of Use: The tool was used exclusively for English language editing, proofreading, and improving the phrasing and grammatical correctness of the manuscript. It was not used to generate any new scientific content, ideas, or data analysis. The authors confirm that they have critically reviewed and edited any AI-generated content and take full responsibility for the integrity, accuracy, and originality of the publication. The authors certify that the original human contribution is maintained and that AI-assisted tools are not listed or cited as authors.

References

[1] Hong S, Kim S, Kim K, Lee H. Clinical approaches for mitochondrial diseases. Cells 2023;12(20):2494. **[Crossref]**

[2] Grier J, Hirano M, Karaa A, Shepard E, Thompson JLP. Diagnostic odyssey of patients with mitochondrial disease: results of a survey. Neurol Genet 2018;4(2):e230. **[Crossref]**

[3] Wen H, Deng H, Li B, et al. Mitochondrial diseases: from molecular mechanisms to therapeutic advances. Signal Transduct Target Ther 2025;10(1):9. **[Crossref]**

[4] Forny P, Footitt E, Davison JE, et al. Diagnosing mitochondrial disorders remains challenging in the omics era. Neurol Genet 2021;7(3):e597. **[Crossref]**

[5] Morava E, van den Heuvel L, Hol F, et al. Mitochondrial disease criteria: diagnostic applications in children. Neurology 2006;67(10):1823-6. **[Crossref]**

[6] Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17(5):405-24. **[Crossref]**

[7] Di Donato S. Multisystem manifestations of mitochondrial disorders. J Neurol 2009;256(5):693-710. **[Crossref]**

[8] Klopstock T, Priglinger C, Yilmaz A, Kornblum C, Distelmaier F, Prokisch H. Mitochondrial disorders. Dtsch Arztebl Int 2021;118(44):741-8. **[Crossref]**

[9] Rouzier C, Pion E, Chaussenot A, et al. Primary mitochondrial disorders and mimics: Insights from a large French cohort. Ann Clin Transl Neurol 2024;11(6):1478-91. **[Crossref]**

[10] van der Ven AT, Johannsen J, Kortüm F, et al. Prevalence and clinical prediction of mitochondrial disorders in a large neuropediatric cohort. Clin Genet 2021;100(6):766-70. **[Crossref]**

[11] Kerr M, Hume S, Omar F, et al. MITO-FIND: a study in 390 patients to determine a diagnostic strategy for mitochondrial disease. Mol Genet Metab 2020;131(1-2):66-82. **[Crossref]**

[12] Schon KR, Horvath R, Wei W, et al. Use of whole genome sequencing to determine genetic basis of suspected mitochondrial disorders: cohort study. BMJ 2021;375:e066288. **[Crossref]**

[13] Grigalionienė K, Burnytė B, Ambrozaitytė L, Utkus A. Wide diagnostic and genotypic spectrum in patients with suspected mitochondrial disease. Orphanet J Rare Dis 2023;18(1):307. **[Crossref]**

[14] Gorman E, Dai H, Feng Y, et al. Experiences from dual genome next-generation sequencing panel testing for mitochondrial disorders: a comprehensive molecular diagnosis. Front Genet 2025;16:1488956. **[Crossref]**

[15] Mahmud S, Biswas S, Afrose S, et al. Use of next-generation sequencing for identifying mitochondrial disorders. Curr Issues Mol Biol 2022;44(3):1127-48. **[Crossref]**

[16] Olimpio C, Paramonov I, Matalonga L, et al. Increased diagnostic yield by reanalysis of whole exome sequencing data in mitochondrial disease. J Neuromuscul Dis 2024;11(4):767-75. **[Crossref]**

acta medica

RESEARCH ARTICLE

Comparative prognostic performance of ELN 2022 and ELN 2024 risk classifications in a Turkish cohort of acute myeloid leukemia patients receiving hypomethylating agents and BCL-2 inhibitors

Selin Küçükyurt Kaya[1], Oğuzhan Koca[2], Lale Aydın Kaynar[1], Emine Merve Savaş[1], Onurcan Azaklı[1], Şahika Zeynep Akı[1], Murat Albayrak[1], Hacer Berna Afacan Öztürk[1], Haktan Bağış Erdem[3], Ahmet Kürşad Güneş[1]

[1]Department of Hematology, Ankara Etlik City Hospital, Ankara, Türkiye
[2]Department of Internal Medicine, Cerrahpaşa Faculty of Medicine, Istanbul University-Cerrahpaşa, İstanbul, Türkiye
[3]Department of Medical Genetics, Ankara Etlik City Hospital, Ankara, Türkiye

Abstract

Objective: The European LeukemiaNet (ELN) 2022 risk classification for acute myeloid leukemia (AML) was primarily developed in cohorts treated with intensive chemotherapy and has demonstrated limited prognostic discrimination in AML patients receiving less-intensive regimens. The recently proposed ELN 2024 classification aims to refine risk stratification in patients treated with less-intensive regimens. We compared the prognostic performance of ELN 2022 and ELN 2024 in a real-world cohort of unfit AML patients treated with azacitidine plus venetoclax.

Materials and Methods: In this retrospective single-center study, 39 newly diagnosed AML patients treated with first-line azacitidine and venetoclax between January 2023 and September 2025 were included. Patients were stratified according to ELN 2022 and ELN 2024 criteria. Overall survival (OS) was analyzed using Kaplan–Meier estimates, log-rank tests, Cox regression, and Harrell's concordance index (C-index).

Results: Median age was 70 years (range, 60–84). Secondary AML was present in 33.3%, and 30.8% harbored TP53 mutations. Under ELN 2022, 64.1% of patients were classified as adverse risk compared with 30.8% under ELN 2024. ELN 2022 did not significantly stratify OS in either three-group or dichotomized analyses (p=0.265 and p=0.199, respectively). In contrast, dichotomized ELN 2024 demonstrated significant survival separation (p=0.041). Adverse risk according to ELN 2024 was associated with inferior OS (HR 2.41, 95% CI 0.98–5.94; p=0.057). The highest discriminatory capacity was observed with the dichotomized ELN 2024 model (C-index 0.697; p=0.021).

Conclusion: In AML patients treated with hypomethylating agents plus venetoclax, ELN 2024 provides improved prognostic discrimination compared with ELN 2022. These findings support the clinical relevance of treatment-context–specific risk stratification in the venetoclax era.

Keywords: acute myeloid leukemia, ELN 2024, genetic risk stratification, venetoclax, hypomethylating agents

Corresponding author: Selin Küçükyurt Kaya • **Email:** dr.skucukyurt@hotmail.com

Received: February 14, 2026 **Accepted:** May 29, 2026 **Published online:** June 28, 2026

Introduction

Acute myeloid leukemia (AML) is a biologically heterogeneous malignancy with historically poor outcomes in older or medically unfit patients who are ineligible for intensive induction chemotherapy. Prior to 2018, treatment options for this population were largely limited to hypomethylating agents (HMAs) or low-dose cytarabine, with median overall survival (OS) rarely exceeding one year [1]. The U.S. Food and Drug Administration's approval of the BCL-2 inhibitor venetoclax in combination with azacitidine or decitabine in 2018 marked a pivotal turning point in AML therapy, leading to substantially higher response rates and improved survival in patients previously considered to have a dismal prognosis [2].

In parallel with therapeutic advances, rapid progress in genomic profiling has fundamentally reshaped risk assessment in AML. Molecular and cytogenetic abnormalities are currently recognized as key determinants of treatment response and survival, forming the backbone of European LeukemiaNet (ELN) risk classifications [3,4]. However, the ELN 2017 and subsequently ELN 2022 recommendations were derived predominantly from cohorts treated with intensive chemotherapy and were not designed to stratify outcomes in patients receiving less-intensive, HMA-based regimens [4,5]. As a result, validation studies in venetoclax-treated or HMA-treated older patients demonstrated suboptimal prognostic discrimination, with a disproportionate number of patients being assigned to the adverse-risk category [6].

These limitations prompted the development of the ELN 2024 genetic risk classification for patients receiving less-intensive therapies, which represents a conceptual shift from therapy-agnostic to treatment-context–specific prognostication. The ELN 2024 framework integrates emerging real-world and clinical trial data from patients treated with HMA monotherapy, HMA plus venetoclax, or azacitidine plus targeted agents such as ivosidenib. Importantly, it emphasizes the dominant adverse prognostic impact of TP53 mutations, while identifying favorable-risk subgroups such as DDX41-mutated or selected NPM1- and IDH-mutated AML, particularly in the absence of activating signaling mutations [6].

Given the widespread adoption of venetoclax-based regimens in real-world clinical practice, there remains a critical unmet need to determine whether the ELN 2024 classification provides superior prognostic stratification compared with ELN 2022 in this specific therapeutic setting. Accordingly, in our study, we aimed to compare the prognostic performance of the ELN 2022 and ELN 2024 risk classifications in patients with AML treated with a combination of HMAs and BCL-2 inhibitors, and to evaluate their ability to discriminate survival outcomes within a real-world cohort.

Materias and Methods

Study design and patient population

This retrospective, single-center observational cohort study included consecutive adult patients (≥18 years) with newly diagnosed AML according to the 2022 World Health Organization (WHO) and International Consensus Classification (ICC) criteria between January 2023 and September 2025 [7,8].

All patients received first-line therapy with azacitidine in combination with venetoclax at the Department of Hematology, Ankara Etlik City Hospital, and those who received at least one dose of both agents were included in the analysis.

The study was approved by the Ankara Etlik City Hospital Ethics Committee (Date: 28-05-2025; Approval No: AEŞH-BADEK1-2025-086) and conducted in accordance with the Declaration of Helsinki.

Data collection and molecular analyses

Demographic, clinical, and laboratory data were retrieved from electronic and paper-based medical records. Collected variables included age at diagnosis, sex, comorbidities, prior azacitidine exposure, cytogenetic findings, and molecular abnormalities at diagnosis.

Conventional cytogenetic analysis was performed using G-banding. Fluorescence in situ hybridization (FISH) was conducted to detect recurrent chromosomal abnormalities, including del(5q), del(7q), +8, del(20q), t(8;21), t(15;17), inv(16), t(9;22), KMT2A rearrangements, DEK/NUP214, and inv(3).

Molecular Profiling and Next-Generation Sequencing (NGS)

Comprehensive somatic genomic profiling was performed using a high-throughput targeted NGS approach designed to evaluate 74 genes frequently mutated in myeloid malignancies and bone marrow failure syndromes. The panel targeted the complete coding regions and essential splice sites of a broad gene set, including NPM1, FLT3 (ITD and TKD), DNMT3A, TET2, TP53, ASXL1, RUNX1, IDH1/2, as well as markers for telomere maintenance (TERC, TERT) and ribosome biogenesis (SBDS, RPL23).

Genomic DNA was isolated from bone marrow aspirates or peripheral blood samples obtained at diagnosis. Library preparation was performed using the SOPHiA™ Myeloid Custom Solution (SOPHiA GENETICS, Saint-Sulpice, Switzerland), employing a hybrid-capture-based target-enrichment methodology. This approach was selected for its superior capacity to provide uniform coverage and detection internal tandem duplications. Sequencing was performed on the Illumina NextSeq 2000 (Illumina, San Diego, CA, USA) platform.

Somatic variants were interpreted and categorized according to the Joint Consensus Recommendations of the Association for Molecular Pathology (AMP), American Society of Clinical Oncology (ASCO), and College of American Pathologists (CAP). A variant allele frequency (VAF) threshold of ≥5% was utilized for clinical risk stratification. However, variants with lower VAFs were meticulously evaluated for key driver genes, such as TP53, where subclonal mutations may carry independent prognostic weight. Based on cytogenetic and molecular findings at diagnosis, patients were stratified according to both the ELN 2022 and ELN 2024 risk classifications.

Endpoints

The primary endpoint of the study was OS, defined as the time from AML diagnosis to death from any cause. Patients who were alive at the time of last follow-up were censored accordingly.

Statistical analysis

All statistical analyses were performed using IBM SPSS Statistics version 25.0 (IBM Corp., Armonk, NY, USA). Categorical variables were expressed as frequencies and percentages. Continuous variables were reported as median and range.

Survival probabilities were estimated using the Kaplan–Meier method and compared between groups using the log-rank test. Cox proportional hazards regression analysis was performed to estimate hazard ratios (HRs) and 95% confidence intervals (CIs). Prognostic discrimination was evaluated using Harrell's concordance index (C-index). To enhance statistical robustness given the small sample size, additional dichotomized analyses (favorable/intermediate vs adverse risk) were performed for both ELN 2022 and ELN 2024. A two-sided p-value <0.05 was considered statistically significant.

Results

A total of 39 patients were included in the study. The median age at diagnosis was 70 years (range, 60–84), and 56.4% of the cohort were male. Baseline demographic, clinical, and molecular characteristics were summarized in Table 1. The most frequently mutated genes identified by NGS were NPM1 (32.4%), DNMT3A (23.5%), TET2 (23.5%), SRSF2 (23.5%), FLT3 (20.6%), ASXL1 (20.6%), and TP53 (20.6%). When patients harboring TP53 mutations detected by Sanger sequencing were additionally included, the overall prevalence of TP53 mutations in the entire cohort increased to 30.8%. Secondary AML accounted for 33.3% of cases, and 20.5% of patients had a documented history of prior azacitidine exposure.

The median follow-up was 7 months (IQR, 3–11). During follow-up, 20 deaths were observed, providing the basis for the OS analyses. Kaplan–Meier analysis of OS for the entire cohort is shown in Figure 1, yielding an estimated 12-month OS rate of 41.1%.

The redistribution of risk categories between ELN 2022 and ELN 2024 was summarized in Figure 2. Under ELN 2022, 3 patients were classified as favorable, 11 as intermediate, and 25 as adverse risk. In contrast, ELN 2024 categorized 17 patients as favorable, 10 as intermediate, and 12 as adverse risk. All patients classified as favorable by ELN 2022 remained favorable under ELN 2024. Among the 11 patients classified as intermediate risk according to ELN 2022, 8 (72.7%) remained in the intermediate category, whereas 3 (27.3%) were reclassified as favorable under ELN 2024. Importantly, among the 25 patients categorized as adverse risk by ELN 2022, only 12 (48%) remained adverse under ELN 2024, while 2 (8%) were reassigned

Table 1. Baseline demographic, clinical, and molecular characteristics of the study cohort

Characteristics	Entire Cohort (n: 39)
Age at diagnosis, years	
Median (Range)	70 (60 – 84)
Age categories, n (%)	
60-69 years	19 (48.7)
70-79 years	17 (43.6)
80-85 years	3 (7.7)
Sex, n (%)	
Female	17 (43.6)
Male	22 (56.4)
ELN 2022 risk classification, n (%)	
Favorable	3 (7.7)
Intermediate	11 (28.2)
Adverse	25 (64.1)
ELN 2024 risk classification, n (%)	
Favorable	17 (43.6)
Intermediate	10 (25.6)
Adverse	12 (30.8)
Prior exposure to azacitidine, n (%)	**8 (20.5)**
Secondary AML*, n (%)	**13 (33.3)**
Recurrent gene mutations detected by NGS, n (%) (N: 34)**	
NPM1	11 (32.4)
FLT3	7 (20.6)
DNMT3A	8 (23.5)
TP53	7** (20.6)
IDH 1 – IDH 2	5 (14.7)
KRAS – NRAS	2 (5.9)
RUNX1	4 (11.8)
TET2	8 (23.5)
CEBPA	4 (11.8)
ASXL1	7 (20.6)
PTPN11	3 (8.8)
SRSF2	8 (23.5)
BCOR	5 (14.7)
Others (ATM, DDX41, JAK2, RAD21, SETBP1, SF3B1, STAG2, U2AF1)	10 (29.4)
Allogeneic HSCT performed, n (%)	4 (10.3)
Median follow-up, months (IQR)	7 (3 – 11)

(ELN, European LeukemiaNet; HSCT, hematopoietic stem cell transplantation; NGS, next-generation sequencing) *Secondary AML included patients with therapy-related AML and those with a documented history of myelodysplastic syndrome or chronic myelomonocytic leukemia prior to AML diagnosis. **NGS-based molecular profiling was performed in 34 patients. Mutation frequencies were calculated exclusively among patients who underwent NGS testing. The TP53 mutation rate presented in the table includes only mutations identified by NGS.

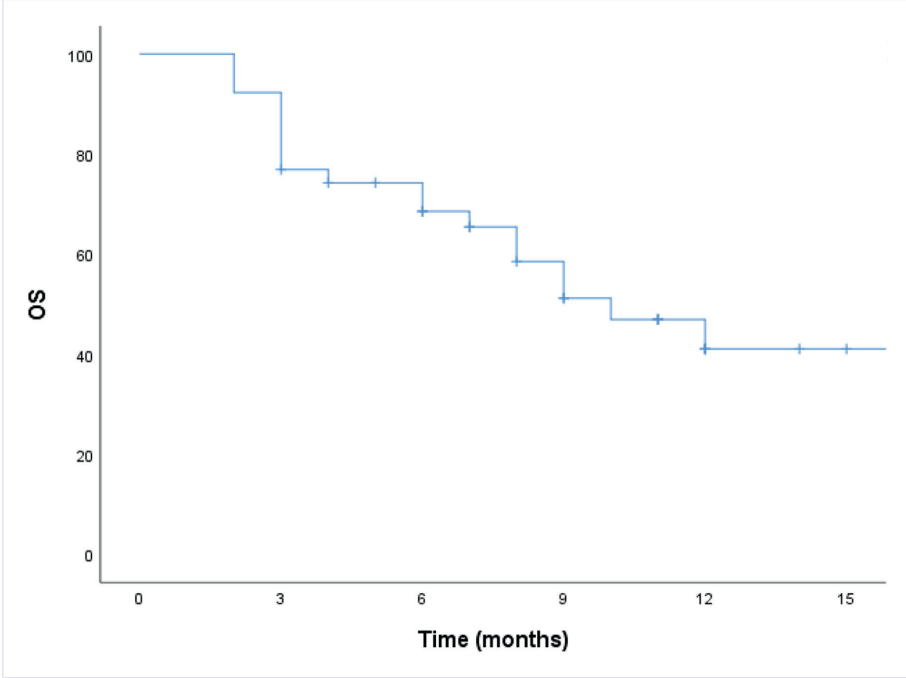


Figure 1. Overall survival curve for the entire cohort (OS, overall survival)

to the intermediate category and 11 (44%) were reclassified as favorable (Figure 2).

Kaplan–Meier survival analyses were performed according to both ELN 2022 and ELN 2024 risk classifications. Using the three-group ELN 2022 model (favorable, intermediate, adverse), no statistically significant difference in OS was observed among risk categories (log-rank p=0.265; Figure 3A). Similarly, when patients were dichotomized as favorable/intermediate versus adverse risk, ELN 2022 did not significantly stratify OS (log-rank p=0.199; Figure 3B). In contrast, ELN 2024 demonstrated improved risk discrimination. Although the three-group analysis did not reach statistical significance (log-rank p=0.125; Figure 4A), dichotomizing into favorable/intermediate versus adverse risk groups resulted in significant separation of the survival curves (log-rank p=0.041; Figure 4B).

In Cox regression analysis, the adverse risk category according to ELN 2022 was not significantly associated with inferior OS compared with the favorable/intermediate risk categories (hazard ratio [HR] 1.98, 95% confidence interval [CI] 0.66–5.97; p=0.225). By contrast, the adverse risk category defined by ELN 2024 showed a stronger association with inferior OS (HR 2.41, 95% CI 0.98–5.94; p=0.057), demonstrating a near-significant trend toward worse survival.

Model discrimination was assessed using Harrell's concordance index (C-index). The ELN 2022 model yielded a C-index of 0.613 (p=0.215; 95% CI, 0.434–0.792). A similar value was observed for the dichotomized ELN 2022 model (C-index 0.612; p=0.221). ELN 2024 showed improved discrimination with a C-index of 0.672 (p=0.051; 95% CI, 0.499–0.845). Notably, the highest discriminative performance was achieved with the dichotomized ELN 2024 model (C-index 0.697; p=0.021; 95% CI, 0.529–0.865), indicating superior prognostic accuracy compared with ELN 2022 (Supplementary Table 1).

When patients with prior azacitidine exposure or those who underwent allogeneic transplantation were excluded, the cohort size decreased; however, the overall direction and magnitude of the survival differences did not change materially (data not shown).

Discussion

In our real-world cohort of AML patients uniformly treated with azacitidine plus venetoclax, ELN 2024 showed improved prognostic discrimination compared

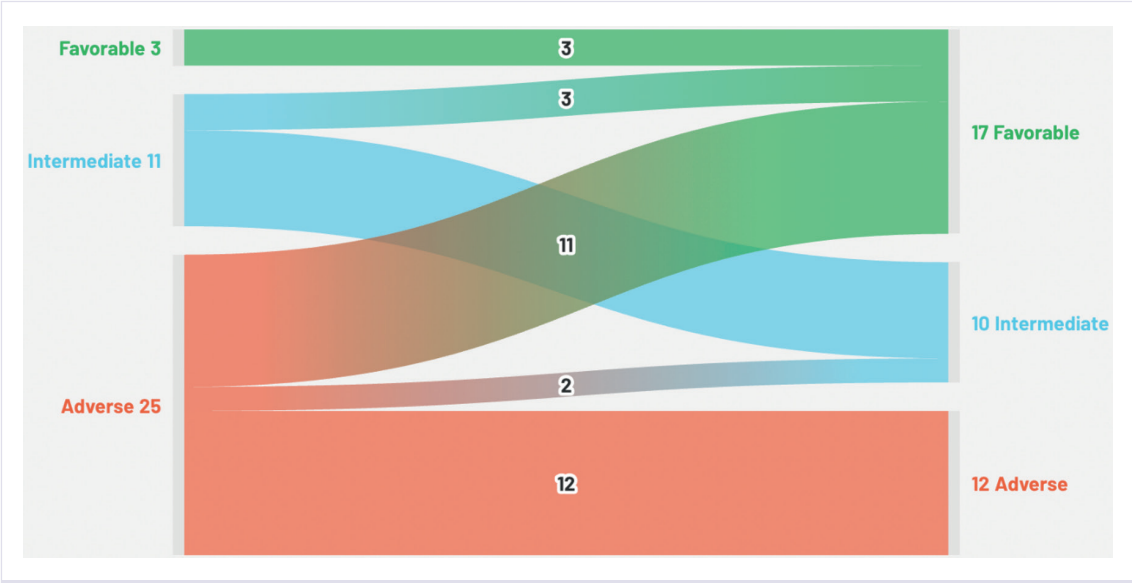


Figure 2. Risk category redistribution between ELN 2022 and ELN 2024 classifications

Graphical representation of patient reclassification across risk categories when transitioning from ELN 2022 to ELN 2024 criteria. Under ELN 2022, 3 patients were classified as favorable, 11 as intermediate, and 25 as adverse risk. In contrast, ELN 2024 categorized 17 patients as favorable, 10 as intermediate, and 12 as adverse risk. All patients classified as favorable according to ELN 2022 remained favorable under ELN 2024. While most intermediate-risk patients retained their classification, a subset was reclassified as favorable. Notably, among the 25 patients categorized as adverse risk by ELN 2022, only 12 remained adverse under ELN 2024, with the remainder reassigned to lower-risk categories.

with ELN 2022. Although three-group comparisons did not consistently reach statistical significance—likely due to the limited sample size—the directionality of results across Kaplan–Meier analyses, Cox regression, and concordance indices generally favored ELN 2024. Notably, the dichotomized ELN 2024 model (favorable/intermediate vs adverse) achieved significant survival separation and the highest discriminative capacity (C-index 0.697), providing preliminary support for its potential clinical utility in the venetoclax era.

AML outcomes vary substantially according to recurrent cytogenetic and molecular abnormalities. The ELN 2017 and 2022 classifications stratify patients into favorable, intermediate, and adverse risk groups based on these features. However, the derivation cohorts primarily consisted of younger patients treated with intensive chemotherapy, with or without allogeneic hematopoietic stem cell transplantation [4,5]. Consequently, ELN 2022 was not designed for patients receiving less-intensive regimens such as HMAs combined with venetoclax. Subsequent validation studies in older or unfit AML patients treated with HMAs with or without venetoclax demonstrated limited prognostic discrimination and an overrepresentation of patients within the adverse-risk category [4,6,9,10]. Such skewed allocation may reduce the clinical utility of ELN-based stratification as treatment paradigms shift toward venetoclax-based combinations in non-intensive settings.

In our cohort, ELN 2022 classified 64.1% of patients as adverse risk, whereas ELN 2024 reduced this proportion to 30.8%, with substantial redistribution to favorable and intermediate categories. This shift reflects the transition from therapy-agnostic risk assessment to treatment-context–specific stratification. By incorporating outcome data from HMA-based regimens, including HMA plus venetoclax and azacitidine combined with targeted agents, ELN 2024 recalibrates the prognostic weight of molecular abnormalities in the non-intensive setting [6,11]-13]. In particular, it emphasizes the dominant adverse impact of TP53 mutations while refining the classification of other molecular subgroups.

The biological rationale is supported by evidence showing that venetoclax-based regimens improve outcomes in NPM1-mutated AML—traditionally a favorable-risk subgroup—whereas TP53-mutated disease continues to carry a poor prognosis despite combination therapy],14]-16]. In our cohort, TP53

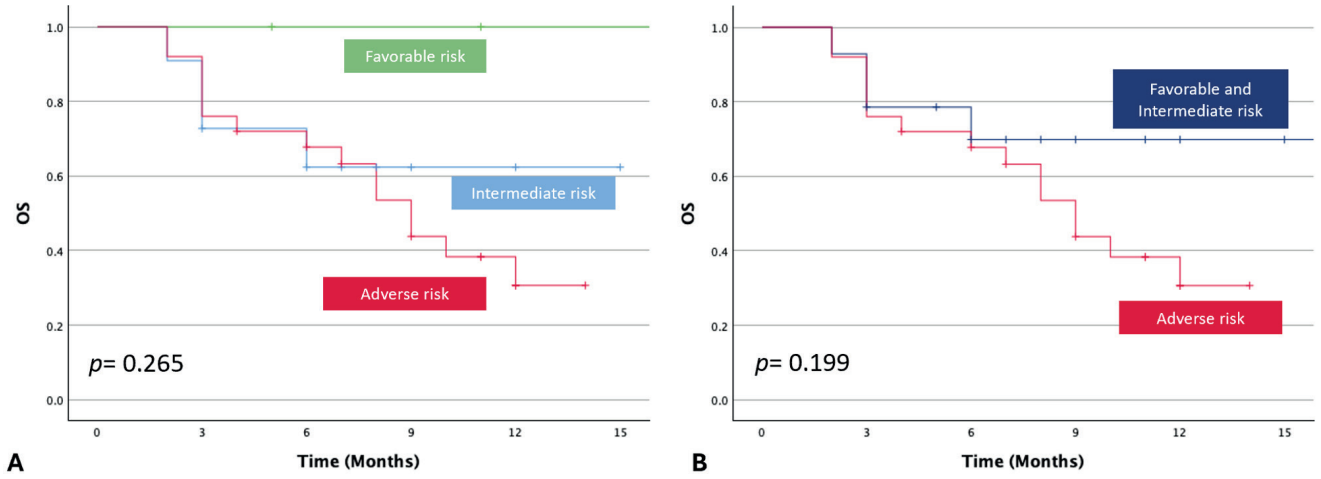


Figure 3. Overall survival according to ELN 2022 risk classification

Kaplan–Meier curves illustrating overall survival (OS) stratified by ELN 2022 risk categories. (A) A three-group analysis (favorable, intermediate, and adverse risk) showed no statistically significant difference in OS among risk groups (log-rank p=0.265). (B) Dichotomized analysis comparing favorable/intermediate versus adverse risk also failed to show significant survival discrimination (log-rank p=0.199).

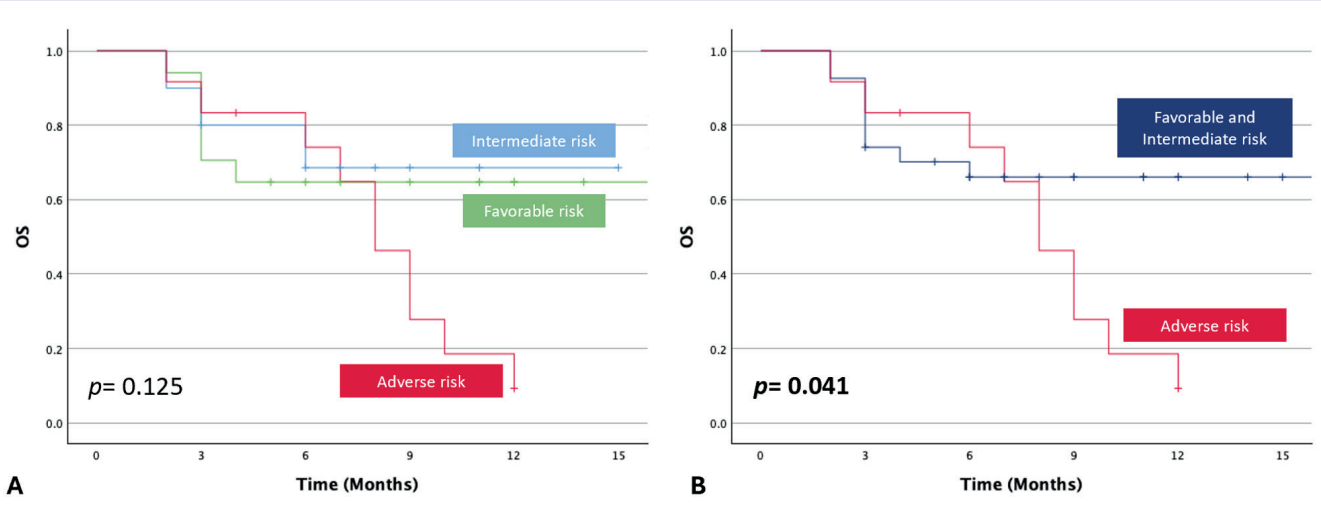


Figure 4. Overall survival according to ELN 2024 risk classification

Kaplan–Meier curves illustrating overall survival (OS) stratified by ELN 2024 risk categories. (A) Three-group analysis (favorable, intermediate, and adverse risk) did not demonstrate a statistically significant difference in OS (log-rank p=0.125). (B) Dichotomized analysis comparing favorable/intermediate versus adverse risk groups showed significant separation of survival curves (log-rank p=0.041), indicating improved prognostic discrimination with ELN 2024.

mutations were present in 30.8% of patients, reflecting the high-risk molecular profile typical of older AML populations treated with non-intensive regimens. The improved discrimination observed with ELN 2024 may be related to more accurate risk allocation of these molecularly defined subgroups within the venetoclax-based treatment context [6].

When comparing our cohort with the pivotal VIALE-A trial, several differences become apparent [14]. In VIALE-A, 286 of 431 patients were assigned to azacitidine–venetoclax, with a median age of 76 years and a median OS of 14.7 months after 20.5 months of follow-up [14]. Prior exposure to HMAs, venetoclax, or chemotherapy for myelodysplastic syndrome was

an exclusion criterion [14]. In contrast, 20.5% of our patients had prior azacitidine exposure, reflecting a less selected real-world population. The proportion of secondary AML was higher in our cohort (33.3% vs 25% in VIALE-A), and TP53 mutation frequency was also increased (30.8% vs 23% in the azacitidine–venetoclax arm) [14]. These adverse biological features likely contributed to the comparatively shorter survival observed in our study. Additionally, real-world factors—including a greater comorbidity burden, variability in supportive care, and treatment interruptions—may further explain differences in outcomes compared with a controlled trial setting. Together, these elements indicate that our cohort represents a more complex, higher-risk population, underscoring the importance of validating prognostic models in routine clinical practice.

Our study has several limitations that should be considered. The modest sample size and short median follow-up (7 months) restrict evaluation of long-term survival and preclude fully adjusted multivariable modeling. In addition, the single-center design, institutional testing practices, referral patterns, and transplant selection strategies may limit the generalizability of our findings. Nonetheless, the consistent directionality of survival separation and effect estimates across analytic methods supports the biological plausibility of ELN 2024 stratification in this context. Larger multicenter studies with longer follow-up are warranted to confirm these findings and further refine prognostic models, potentially integrating molecular risk with measurable residual disease (MRD) assessment to enhance risk prediction in the venetoclax era [17].

Conclusion

In this real-world Turkish AML cohort treated with azacitidine plus venetoclax, the ELN 2024 risk classification showed improved prognostic discrimination compared with ELN 2022. ELN 2024 more accurately redistributed patients across risk categories and achieved superior survival separation and concordance performance. These findings provide preliminary support for the potential clinical utility of treatment-context–specific risk stratification and reinforce the need for continued external validation of evolving genetic classification systems in contemporary AML practice. With larger patient cohorts and extended follow-up, future studies may enable the development and refinement of novel molecularly driven prognostic models tailored specifically to venetoclax-based treatment settings, potentially integrating genomic, clinical, and MRD parameters to further enhance risk prediction.

Author contributions

Conception: S.K.K., A.K.G.; Design: S.K.K., L.A.K., E.M.S., Ş.Z.A., A.K.G.; Data acquisition: S.K.K., L.A.K., E.M.S., O.A., Ş.Z.A., M.A., H.B.A.Ö., H.B.E., A.K.G.; Data analysis: S.K.K., O.K., Ş.Z.A., A.K.G.; Data interpretation: S.K.K., O.K., A.K.G.; Drafting of the manuscript: S.K.K., O.K., O.A., Ş.Z.A., M.A., H.B.A.Ö., H.B.E., A.K.G.; Critical revision of the manuscript: S.K.K., O.K., L.A.K., E.M.S., Ş.Z.A., A.K.G. All authors reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of this study.

Ethical approval

This study was approved by the Ankara Etlik City Hospital Ethics Committee (Date: May 28, 2025, Decision/Protocol No: AEŞH-BADEK1-2025-086). Informed consent was obtained from all participants involved in this study.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The authors declare that this study received no funding.

Generative AI statement

The authors declare that no generative AI or AI-assisted technologies were used in the writing or preparation of this study.

References

[1] Kucukyurt S, Eskazan AE. New drugs approved for acute myeloid leukaemia in 2018. Br J Clin Pharmacol 2019;85(12):2689-93. **[Crossref]**

[2] Pratz KW, Jonas BA, Pullarkat V, et al. Long-term follow-up of VIALE-A: venetoclax and azacitidine in chemotherapy-ineligible untreated acute myeloid leukemia. Am J Hematol 2024;99(4):615-24. **[Crossref]**

[3] Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med 2016;374(23):2209-21. **[Crossref]**

[4] Döhner H, Wei AH, Appelbaum FR, et al. Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. Blood 2022;140(12):1345-77. **[Crossref]**

[5] Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2017;129(4):424-47. **[Crossref]**

[6] Döhner H, DiNardo CD, Appelbaum FR, et al. Genetic risk classification for adults with AML receiving less-intensive therapies: the 2024 ELN recommendations. Blood 2024;144(21):2169-73. **[Crossref]**

[7] Arber DA, Orazi A, Hasserjian RP, et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. Blood 2022;140(11):1200-28. **[Crossref]**

[8] Khoury JD, Solary E, Abla O, Akkari YMN. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: myeloid and histiocytic/dendritic neoplasms. Leukemia 2022;36(7):1703-19. **[Crossref]**

[9] Jahn E, Saadati M, Fenaux P, et al. Correction: clinical impact of the genomic landscape and leukemogenic trajectories in non-intensively treated elderly acute myeloid leukemia patients. Leukemia 2023;37(11):2336-37. **[Crossref]**

[10] Döhner H, Pratz KW, DiNardo CD, et al. Genetic risk stratification and outcomes among treatment-naive patients with AML treated with venetoclax and azacitidine. Blood 2024;144(21):2211-22. **[Crossref]**

[11] Döhner H, Dolnik A, Tang L, et al. Cytogenetics and gene mutations influence survival in older patients with acute myeloid leukemia treated with azacitidine or conventional care. Leukemia 2018;32(12):2546-57. **[Crossref]**

[12] Montesinos P, Recher C, Vives S, et al. Ivosidenib and azacitidine in IDH1-mutated acute myeloid leukemia. N Engl J Med 2022;386(16):1519-31. **[Crossref]**

[13] Pollyea DA, Pratz KW, Wei AH, et al. Outcomes in patients with poor-risk cytogenetics with or without tp53 mutations treated with venetoclax and azacitidine. Clin Cancer Res 2022;28(24):5272-9. **[Crossref]**

[14] DiNardo CD, Jonas BA, Pullarkat V, et al. Azacitidine and venetoclax in previously untreated acute myeloid leukemia. N Engl J Med 2020;383(7):617-29. **[Crossref]**

[15] Lachowiez CA, Loghavi S, Kadia TM, et al. Outcomes of older patients with NPM1-mutated AML: current treatments and the promise of venetoclax-based regimens. Blood Adv 2020;4(7):1311-20. **[Crossref]**

[16] Tiong IS, Dillon R, Ivey A, et al. Venetoclax induces rapid elimination of NPM1 mutant measurable residual disease in combination with low-intensity chemotherapy in acute myeloid leukaemia. Br J Haematol 2021;192(6):1026-30. **[Crossref]**

[17] Wei AH, Iland HJ, DiNardo CD, Reynolds J. Measurable residual disease intervention in AML: a new therapeutic horizon. Blood 2026;147(1):13-23. **[Crossref]**

acta medica

RESEARCH ARTICLE

Biased and inadequate, or trustworthy and sufficient? Evaluating the YouTube videos on platelet-rich plasma in orthopedics

Gökhan Ayık[1], Orhan Mete Karademir[2], Ulaş Can Kolaç[1], Erdi Özdemir[3], Gazi Huri[4]

[1]Department of Orthopedics and Traumatology, Hacettepe University, Ankara, Türkiye
[2]Faculty of Medicine, Hacettepe University, Ankara, Türkiye
[3]Department of Orthopedics, School of Medicine, The University of Alabama at Birmingham, Alabama, USA
[4]Department of Orthopaedics and Traumatology, Aspetar, FIFA Medical Center of Excellence, Doha, Qatar

Abstract

Objective: Social media has transformed patient-physician interactions, with YouTube emerging as a major contributor. Platelet-rich plasma (PRP) therapy has gained substantial popularity in orthopedics despite ongoing debate regarding its effectiveness. As orthobiologic treatments are increasingly promoted directly to consumers, concerns have emerged regarding the quality, reliability, and potential bias of online information. This study aimed to evaluate the quality, reliability, and bias of YouTube videos related to PRP in orthopedics.

Materials and Methods: A cross-sectional analysis of YouTube was performed, using the search terms "prp," "prp knee," and "prp shoulder." The first 50 results were recorded, and after exclusions, eligible videos were analyzed. Video metadata were recorded and content sources were categorized as physician, medical institution, or non-medical sources. Information quality was assessed using Brief DISCERN instrument, JAMA Criteria, and Global Quality Score (GQS). Bias was evaluated based on predefined criteria including promotional language and testimonial-driven claims.

Results: A total of 111 videos were analyzed. Physicians produced 49.5% of videos, followed by institutions (39.6%) and non-medical sources (10.8%). Overall information quality was low-to-moderate, with median Brief DISCERN, JAMA, and GQS scores of 17, 3, and 3, respectively. Significant differences were observed between uploader types for all quality metrics (DISCERN $p=0.008$; JAMA and GQS $p<0.001$). Physicians demonstrated higher quality compared to medical institutions and non-medical sources. Overall, 40.5% of videos were classified as biased, with no significant association between source and bias ($p=0.0516$). Non-biased videos had higher JAMA ($p=0.038$) and GQS ($p=0.025$) scores. Longer videos were associated with higher quality and engagement (all $p<0.001$), while popularity metrics were not associated with information quality.

Conclusion: YouTube videos on PRP in orthopedics demonstrate variable quality, incomplete transparency, and a notable proportion of potentially biased content. Physician-generated content is associated with higher quality, although overall reliability and transparency remain inconsistent. These findings highlight both the need and the opportunity for more accurate, transparent, balanced, and evidence-based content to better support patient decision-making.

Keywords: platelet-rich plasma, orthopedics, social media, health information quality, bias

Introduction

In the current landscape of digital health, social media platforms have substantially transformed the traditional patient-provider dynamic, shifting toward more participatory and information-driven interactions [1-3]. Among these platforms, YouTube has emerged as one of the most visited websites globally and serves

Corresponding author: Gökhan Ayık • **Email:** drgayik@gmail.com

Received: April 4, 2026 **Accepted:** May 6, 2026 **Published online:** June 28, 2026

as a major repository for health-related information [4-6]. Importantly, online video platforms may also provide accessible and practical educational content for patients when information is accurate, transparent, and appropriately contextualized [6-9].

This shift toward digital self-education is particularly evident in orthopedics, where patients frequently access online resources prior to consultation and explore nonoperative treatment options for chronic musculoskeletal conditions [9-12]. These resources may improve familiarity with treatment options, but they may also shape expectations before a clinician has had the opportunity to discuss indications, uncertainties, risks, and alternatives [9,13].

Public interest in PRP has increased in recent years, as reflected by online search trends (Figure 1). The global PRP market has demonstrated rapid growth, with estimates exceeding $600 million by 2025, largely driven by its increasing application in knee osteoarthritis and shoulder pathologies [14,15]. Although numerous clinical trials and meta-analyses suggest that PRP may improve pain and functional outcomes in selected musculoskeletal conditions, its clinical effectiveness remains debated [16-22]. This uncertainty is partly due to substantial variation in preparation methods, composition, and application protocols across studies.

The growing demand for "regenerative" therapies has led to rapid commercialization of orthobiologics, often accompanied by direct-to-consumer marketing strategies that may outpace established evidence-based guidelines [9,23,24]. In this setting, promotional narratives, selective presentation of benefits, and limited discussion of uncertainty may affect patient expectations and informed decision-making [7,25-27].

Previous studies evaluating medical and orthobiologic content on YouTube have shown variable information quality and incomplete transparency [5,28,29]. However, there remains a relative paucity of studies specifically examining PRP-related YouTube videos in orthopedics while assessing both informational quality and potential commercial or promotional bias [29,30]. This combined assessment represents the main contribution of the present study.

Therefore, the aim of this study was to evaluate the quality, reliability, and bias of YouTube videos addressing the use of PRP in orthopedic practice. We hypothesized that PRP-related videos would demonstrate variable information quality and that a substantial proportion would contain potentially promotional or biased content.

Materials and Methods

No ethical approval was required for this study, as it did not involve human participants, animal subjects, or identifiable personal data.

On 27 March 2026, a cross-sectional search of YouTube was performed using the search terms "prp", "prp knee", and "prp shoulder". Searches were performed on a newly configured MacBook (Tahoe 26.4) using Safari Version 26.4, with cookies and tracking disabled, along with a newly created YouTube account based in the United States. Searches were performed using YouTube's default relevance-based sorting, which was

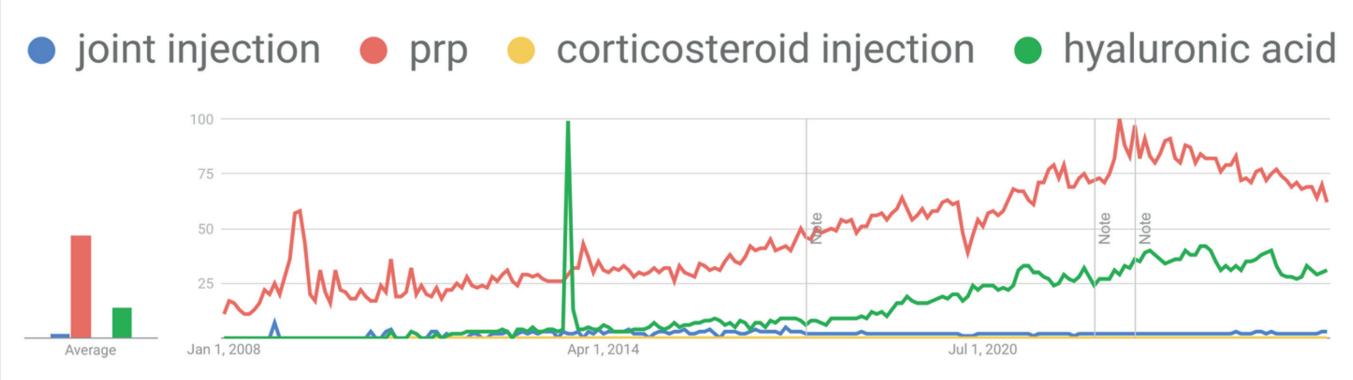


Figure 1. YouTube search trends for joint injection, platelet-rich plasma (PRP), corticosteroid injection, and hyaluronic acid between January 1, 2008, and March 1, 2026, presented on a normalized 0-100 scale

selected to reflect the order most likely encountered by a typical user. The first 50 results for each search term were screened because users are more likely to interact with videos appearing early in search results. Beyond this range, results increasingly included less relevant recommendations. Videos appearing in more than one search were recorded once, and duplicate entries were removed before final analysis. Non-English content, videos not related to orthopedic PRP applications, and inaccessible or duplicate videos were excluded. Metadata were collected for each video, including duration, view count, likes, comments, and video age. The creator of each video was categorized as a physician, medical institution, or non-medical source. To account for the variation in video age and its impact on popularity, a view ratio was calculated as the total number of views divided by the number of days since upload. An interaction ratio was also calculated to assess user engagement, defined as the sum of likes and comments divided by the total view count.

Information quality and reliability were evaluated using three scoring systems: the Brief DISCERN instrument, which comprises six questions scored from 1 to 5, yielding a total score ranging from 6 to 30, the JAMA (Journal of the American Medical Association) Criteria (one point for each of the four criteria), and the Global Quality Score (GQS) (scored on a 1 to 5 Likert scale) (Table 1). Source neutrality was assessed using a binary bias score, with 1 indicating the presence of bias and 0 indicating neutral presentation. Criteria for a bias score of 1 included the use of hyperbolic descriptors (e.g., “miracle solution” or “liquid gold”), the promotion of proprietary medical brands, or a heavy reliance on patient testimonials to suggest guaranteed outcomes. All videos were independently evaluated by two authors. In instances where initial scoring was not identical, discrepancies were resolved through consensus discussion with the senior author.

Statistical analysis was performed using IBM SPSS Statistics V23 (IBM Corp., Armonk, NY, USA). Descriptive data were presented as median, interquartile range (IQR), minimum, maximum, and first and third quartiles for continuous variables, and as frequencies and percentages for categorical variables. The normality of continuous variables was assessed using the Shapiro–Wilk test and histogram analysis. As the data were not normally distributed, nonparametric tests were applied. Comparisons between two groups were performed using the Mann–Whitney U test, while comparisons across multiple groups were conducted using the Kruskal-Wallis test. When significant differences were identified, post hoc pairwise comparisons were performed using Dunn’s test with Bonferroni correction. Categorical variables were analyzed using the chi-square test or Fisher’s exact test, as appropriate. Correlations between continuous variables were assessed using Spearman’s rank correlation coefficient. A p-value of <0.05 was considered statistically significant.

Results

A total of 150 videos were initially identified, and after applying exclusion criteria, 111 videos were included for final analysis. The median video age was 38.06 (IQR:44.13) months, with a median video duration of 3.83 (IQR: 6.38) minutes, and a median view count of 3807 (IQR:25421). The median view ratio was 6.9 (34.61), and the median interaction ratio was 0.0147 (IQR:0.0205).

Regarding content sources, 49.5% of videos were uploaded by physicians, 39.6% by medical institutions, and 10.8% by non-medical sources. The median Brief DISCERN score was 17 (IQR:6), corresponding to low-to-moderate information quality. The median JAMA score was 3 (IQR:1) and median Global Quality Score (GQS) was 3 (IQR:2). According to the predefined criteria, 32.7% of videos from physicians, 54.5% of videos from medical institutions, and 25.0% of videos from non-medical sources were biased, with a total of 40.5% of videos being classified as biased based on predefined criteria (Table 2) (Figure 2).

When grouped by uploader type, a significant difference was observed among physicians, medical institutions, and non-medical sources for DISCERN ($p=0.008$), JAMA ($p<0.001$), and GQS ($p<0.001$) scores. Physicians had significantly higher DISCERN scores compared to medical institutions ($p=0.002$), while no significant difference was observed between physicians and non-medical sources ($p=0.185$) or between medical institutions and non-medical sources ($p=0.534$). For JAMA and GQS scores, physician-generated videos had significantly higher scores than videos uploaded by medical institutions ($p=0.0004$ and $p=0.0002$, respectively) and non-medical sources ($p=0.0071$ and $p=0.0020$, respectively), whereas no significant differences were found between medical institutions and non-medical sources ($p=0.659$ and $p=0.472$, respectively). There was no statistically significant

Table 1. Scoring systems

Question no	**Brief DISCERN question**
1	Is it clear what sources of information were used to compile the publication (other than the author or producer)?
2	Is it clear when the information used or reported in the publication was produced?
3	Does it describe how each treatment works?
4	Does the publication describe the benefits of each treatment?
5	Does it describe the risks of each treatment?
6	Does it describe how the treatment choices affect overall quality of life?
Score	**Global quality description**
1	Low quality, video information flow weak, most information missing, not beneficial for patients.
2	Usually, low quality and low flow of information, some listed information and many important issues are missing, very limited use for patients.
3	Moderate quality, the insufficient flow of information, and some important information is sufficiently discussed, but some are poorly discussed and somewhat useful for patients.
4	Good quality and generally good information flow. Most of the relevant information is listed, but some topics are not covered, useful for patients.
5	Excellent quality and information flow, very useful for patients.
JAMA criteria	**Description**
Authorship	Are the authors, contributors, and their credentials clearly listed?
Attribution	Are all sources of information, including references and data sources, cited?
Disclosure	Are potential conflicts of interest, funding sources, or commercial interests disclosed?
Currency	Is the date of publication or last update clearly shown?

association between creator type and bias, although the difference approached significance (p=0.0516).

Compared to biased videos, non-biased videos showed higher JAMA (p=0.038) and GQS (p=0.025) scores, but no significant differences were found in terms of View Ratio (p=0.087), Interaction ratio (p=0.220), and Brief DISCERN scores (p=0.080). Longer videos showed higher View Ratio (p<0.001), Interaction ratio (p<0.001), Brief DISCERN scores (p<0.001), JAMA scores (p<0.001), and GQS scores (p<0.001).

Discussion

This study found substantial variability in the quality of YouTube videos addressing PRP use in orthopedics. Overall, videos exhibited low-to-moderate information quality and limited adherence to transparency criteria. A notable proportion of videos (40.5%) met the predefined criteria for bias; however, the association between uploader type and bias did not reach statistical significance (p=0.0516). Therefore, these findings should be interpreted as evidence of potentially promotional framing in a meaningful subset of videos rather than definitive proof of widespread commercial bias.

These findings are consistent with previous literature indicating that online health information often lacks completeness, proper sourcing, and accountability [5,28]. Although no significant association was observed between view-based popularity metrics and information quality, longer videos were consistently associated with higher quality scores and greater engagement. This suggests that content depth, rather than popularity, may be a more meaningful indicator of informational value. It also supports prior evidence that views, likes, and engagement are unreliable proxies for medical accuracy [28].

Table 2. Descriptive statistics

Variables		**Median (1st and 3rd quartiles)**	**Minimum - maximum**
Video Age (months)		38.06 (16-60.13)	0.03-148.76
Video Length (minutes)		3.83 (1.93-8.31)	0.43-69.50
View Count		3807 (514-25835)	10-538209
View Ratio		6.90 (0.47-35.08)	0.05-290.80
Like Count		42 (3-316)	0.00-7800
Comment Count		7 (0-53)	0-809
Interaction Ratio		0.0147 (0.0056-0.0261)	0.0000-0.2000
Brief DISCERN Score		17 (15-21)	10-29
JAMA Score		3 (2-3)	1-4
GQS Score		3 (2-4)	1-5
Variables		**Frequencies**	**Percentages**
Creator	Physician	55	49.5%
	Medical Institutions	44	39.6%
	Non-Medical Sources	12	10.8%
Bias	Biased	45	40.5%
	Non-biased	66	59.5%

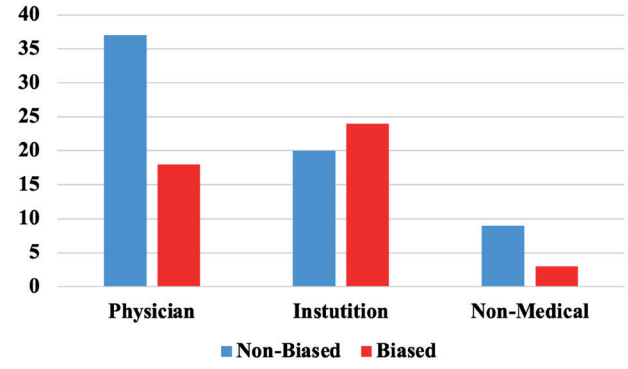


Figure 2. Distribution of biased and non-biased content according to source type

A key strength of this study was the simultaneous evaluation of informational quality and bias. Promotional characteristics, including hyperbolic language, brand-oriented messaging, and testimonial-driven narratives, were observed in a subset of videos. Videos classified as biased demonstrated no significant differences in Brief DISCERN scores, but significantly lower JAMA and GQS scores, which may suggest that commercially driven content may be associated with reduced scientific rigor and limited discussion of uncertainty. This aligns with previous studies highlighting deficiencies in transparency and evidence-based communication in YouTube health content [5]. The influence of commercial bias is particularly relevant in orthobiologics, where direct-to-consumer marketing has expanded rapidly and may shape patient perceptions independent of the underlying evidence [7,23].

Differences according to uploader type further contextualize these findings. Physician-generated videos demonstrated higher quality and reliability scores compared to those produced by non-medical sources, consistent with prior reports that clinician involvement improves the structure and credibility of online medical information [1]. Videos produced by medical institutions were not superior to those uploaded by non-medical sources, suggesting that institutional affiliation alone does not guarantee content quality. However, even physician-produced content did not consistently meet transparency standards, particularly regarding disclosure and referencing, which are essential components of the JAMA criteria [31]. This suggests that while professional involvement improves content quality, it does not fully address issues related to accountability and completeness.

The present findings also align with the broader literature evaluating YouTube as a source of health information. Multiple systematic reviews have concluded that YouTube contains heterogeneous and frequently unreliable medical content, with significant variability across topics and methodologies [5,28]. Within orthopedics, recent studies on orthobiologics have similarly reported low-to-moderate reliability and incomplete information, although some reports focusing specifically on PRP have suggested acceptable quality depending on study design and evaluation criteria [29,30]. These discrepancies likely reflect differences in search strategies, scoring systems, and the rapidly evolving nature of online content [32].

The susceptibility of PRP-related content to oversimplification and promotional framing may be explained by the inherent complexity and heterogeneity of PRP therapy. Although some studies have demonstrated clinical benefits in conditions such as knee osteoarthritis and rotator cuff tendinopathy, others have reported no significant advantage compared to placebo [16,17]. Meta-analyses suggest that outcomes may vary depending on factors such as platelet concentration and preparation methods [18]. This variability is further compounded by the lack of standardized classification systems and reporting protocols, which limits comparability across studies and complicates interpretation [19,22]. As a result, nuanced scientific findings may be reduced to simplified or exaggerated claims in online media, where clear and appealing narratives are often favored over balanced explanations [18].

Commercialization likely also amplifies this phenomenon. The expanding PRP market and increasing demand for regenerative therapies have created strong incentives for direct-to-consumer promotion, often emphasizing accessibility and effectiveness without adequately addressing limitations or uncertainties [15]. Similar patterns have been observed in other areas of regenerative medicine, including stem cell therapies, where misleading or exaggerated claims have raised ethical and regulatory concerns [7,23]. These dynamics contribute to the proliferation of persuasive but potentially misleading content, reinforcing the role of commercial bias in shaping online health information [7].

The clinical implications of these findings are relevant but should be interpreted in light of the study design. Biased or low-quality information may influence patient expectations, contribute to therapeutic misconception, and affect decision-making processes. Informed consent requires a clear understanding of benefits, risks, and alternatives; however, promotional content may distort this balance by emphasizing positive outcomes while minimizing uncertainty [25,27]. Moreover, exposure to online health information has been shown to affect patient behavior, including treatment preferences and healthcare utilization, and may influence the patient-physician relationship [2,26]. In this context, PRP-related online content should be considered an important component of patient education rather than a direct measure of clinical outcomes.

These findings highlight the importance of proactive patient education and the opportunity to improve online resources. Clinicians should routinely inquire about patients' exposure to online information and provide guidance toward reliable, evidence-based resources. Open discussion of online content may enhance patient understanding and strengthen the therapeutic relationship [1,2]. Additionally, healthcare professionals and medical institutions can improve the online information environment by producing concise, balanced, and transparent videos that explain indications, expected benefits, risks, alternatives, uncertainty, and the heterogeneity of PRP preparations in accordance with established quality standards such as the JAMA criteria [31].

Several limitations should be considered. YouTube is a dynamic platform, and search results may vary over time and across users, limiting reproducibility. The cross-sectional design, restriction to English-language videos, and use of predefined search terms may limit generalizability and may not capture the full diversity of available content. Furthermore, while the Brief DISCERN instrument, JAMA criteria, and GQS are widely used, they were not specifically developed for video-based platforms and may not fully reflect all aspects of informational accuracy. The assessment of bias, although based on predefined criteria and independent review, also involves a degree of subjectivity. In addition, formal inter-rater reliability coefficients were not calculated after consensus scoring, which should be addressed in future studies.

Future research should focus on longitudinal analyses to evaluate temporal changes in content quality and incorporate more detailed assessments of scientific

accuracy and evidence levels. Evaluating the impact of online content on patient knowledge, expectations, and clinical decision-making would further enhance understanding of its real-world implications. Additionally, future studies may benefit from more granular analysis of PRP-related content, including whether videos adequately address treatment heterogeneity and reporting standards.

Conclusion

YouTube videos on PRP in orthopedics demonstrate variable quality, incomplete transparency, and a notable proportion of potentially biased content. Physician-generated content is associated with higher quality, although overall reliability and transparency remain inconsistent. These findings underscore the need for more accurate, transparent, balanced, and evidence-based online content to support informed patient decision-making.

Author contributions

Conception: G.A., O.M.K., U.C.K., E.O., G.H.; Design: G.A., O.M.K., U.C.K.; Data acquisition: O.M.K., U.C.K.; Data analysis: O.M.K., E.O.; Data interpretation: G.A., O.M.K., E.O.; Drafting of the manuscript: G.A., O.M.K.; Critical revision of the manuscript: U.C.K., E.O., G.H. All authors reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of this study.

Ethical approval

Ethics committee approval and informed consent were not required for this study.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The authors declare that this study received no funding.

Generative AI statement

The authors declare that no generative AI or AI-assisted technologies were used in the writing or preparation of this study.

References

[1] BBenetoli A, Chen TF, Aslani P. How patients' use of social media impacts their interactions with healthcare professionals. Patient Educ Couns 2018;101(3):439-44. **[Crossref]**

[2] Tan SSL, Goonawardene N. Internet health information seeking and the patient-physician relationship: a systematic review. J Med Internet Res 2017;19(1):e9. **[Crossref]**

[3] Ventola CL. Social media and health care professionals: benefits, risks, and best practices. P T 2014;39(7):491-520.

[4] Monteiro Grilo A, Ferreira AC, Pedro Ramos M, Carolino E, Filipa Pires A, Vieira L. Effectiveness of educational videos on patient's preparation for diagnostic procedures: systematic review and meta-analysis. Prev Med Rep 2022;28:101895. **[Crossref]**

[5] Madathil KC, Rivera-Rodriguez AJ, Greenstein JS, Gramopadhye AK. Healthcare information on YouTube: a systematic review. Health Informatics J 2015;21(3):173-94. **[Crossref]**

[6] Wong DKK, Cheung MK. Online health information seeking and ehealth literacy among patients attending a primary care clinic in hong kong: a cross-sectional survey. J Med Internet Res 2019;21(3):e10831. **[Crossref]**

[7] Erikainen S, Couturier A, Chan S. Marketing experimental stem cell therapies in the UK: biomedical lifestyle products and the promise of regenerative medicine in the digital era. Sci Cult (Lond) 2019;29(2):219-44. **[Crossref]**

[8] Ashkanani H, Asery R, Bokubar F, et al. Web-based health information seeking among students at kuwait university: cross-sectional survey study. JMIR Form Res 2019;3(4):e14327. **[Crossref]**

[9] Meunier A, Posadzy K, Tinghög G, Aspenberg P. Risk preferences and attitudes to surgery in decision making. Acta Orthop 2017;88(5):466-71. **[Crossref]**

[10] Dekkers T, Melles M, Groeneveld BS, de Ridder H. Web-based patient education in orthopedics: systematic review. J Med Internet Res 2018;20(4):e143. **[Crossref]**

[11] Koenig S, Nadarajah V, Smuda MP, Meredith S, Packer JD, Henn RF. Patients' use and perception of internet-based orthopaedic sports medicine resources. Orthop J Sports Med 2018;6(9):2325967118796469. **[Crossref]**

[12] Duymus TM, Karadeniz H, Çaçan MA, et al. Internet and social media usage of orthopaedic patients: a questionnaire-based survey. World J Orthop 2017;8(2):178-186. **[Crossref]**

[13] Kennedy BL, Currie GR, Kania-Richmond A, Emery CA, MacKean G, Marshall DA. Factors that patients consider in their choice of non-surgical management for hip and knee osteoarthritis: formative qualitative research for a discrete choice experiment. Patient 2022;15(5):537-550. **[Crossref]**

[14] Freiberger C, Kale NN, Gallagher ME, Ierulli VK, O'Brien MJ, Mulcahey MK. Patients prefer medical facts and educational videos from sports medicine surgeons on social media. Arthrosc Sports Med Rehabil 2023;5(3):e649-56. **[Crossref]**

[15] Platelet rich plasma market size industry report, 2033. Available at: https://www.grandviewresearch.com/industry-analysis/platelet-rich-plasma-prp-market.

[16] Bennell KL, Paterson KL, Metcalf BR, et al. effect of intra-articular platelet-rich plasma vs placebo injection on pain and medial tibial cartilage volume in patients with knee osteoarthritis: The RESTORE Randomized Clinical Trial. JAMA 2021;326(20):2021-30. **[Crossref]**

[17] Roy M, Reddy MH, Das D, Priyanshu , Chandrakar D, Elavarasu AM. Effectiveness of platelet-rich plasma in treating rotator cuff tendinopathy: a systematic review and meta-analysis. J Orthop Case Rep 2025;15(3):265-74. **[Crossref]**

[18] Bensa A, Previtali D, Sangiorgio A, Boffa A, Salerno M, Filardo G. PRP Injections for the treatment of knee osteoarthritis: the improvement is clinically significant and influenced by platelet concentration: a meta-analysis of randomized controlled trials. Am J Sports Med 2025;53(3):745-754. **[Crossref]**

[19] DeLong JM, Russell RP, Mazzocca AD. Platelet-rich plasma: the PAW classification system. Arthroscopy 2012;28(7):998-1009. **[Crossref]**

[20] Dohan Ehrenfest DM, Rasmusson L, Albrektsson T. Classification of platelet concentrates: from pure platelet-rich plasma (P-PRP) to leucocyte- and platelet-rich fibrin (L-PRF). Trends Biotechnol 2009;27(3):158-67. **[Crossref]**

[21] Rossi LA, Murray IR, Chu CR, Muschler GF, Rodeo SA, Piuzzi NS. Classification systems for platelet-rich plasma. Bone Joint J 2019;101-B(8):891-896. **[Crossref]**

[22] Mautner K, Malanga GA, Smith J, et al. A call for a standard classification system for future biologic research: the rationale for new PRP nomenclature. PM R 2015;7(4 Suppl):S53-9. **[Crossref]**

[23] Kingery MT, Schoof L, Strauss EJ, Bosco JA, Halbrecht J. Online direct-to-consumer advertising of stem cell therapy for musculoskeletal injury and disease: misinformation and violation of ethical and legal advertising parameters. J Bone Joint Surg Am 2020;102(1):2-9. **[Crossref]**

[24] Important patient and consumer information about regenerative medicine therapies 2024. Available at: https://www.fda.gov/vaccines-blood-biologics/consumers-biologics/important-patient-and-consumer-information-about-regenerative-medicine-therapies.

[25] Jansen LA. Informed consent, therapeutic misconception, and unrealistic optimism. Perspect Biol Med 2020;63(2):359-73. **[Crossref]**

[26] Bujnowska-Fedak MM, Węgierek P. The impact of online health information on patient health behaviours and making decisions concerning health. Int J Environ Res Public Health 2020;17(3):880. **[Crossref]**

[27] Borges do Nascimento IJ, Pizarro AB, Almeida JM, et al. Infodemics and health misinformation: a systematic review of reviews. Bull World Health Organ 2022;100(9):544-561. **[Crossref]**

[28] Osman W, Mohamed F, Elhassan M, Shoufan A. Is YouTube a reliable source of health-related information? A systematic review. BMC Med Educ 2022;22(1):382. **[Crossref]**

[29] Sachs JP, Weissman AC, Wagner KR, et al. YouTube is an inconsistent source of information on orthobiologics: implications for content quality, reliability, comprehensiveness, and patient decision making. Arthroscopy 2025;41(10):4225-34. **[Crossref]**

[30] Eravsar E, Mercan N. Platelet-rich plasma videos for knee osteoarthritis on YouTube are of acceptable quality and content. Medicine (Baltimore) 2025;104(30):e43437. **[Crossref]**

[31] Silberg WM, Lundberg GD, Musacchio RA. Assessing, controlling, and assuring the quality of medical information on the Internet: caveant lector et viewor--let the reader and viewer beware. JAMA 1997;277(15):1244-5.

[32] Sui W, Sui A, Rhodes RE. What to watch: practical considerations and strategies for using YouTube for research. Digit Health 2022;8:20552076221123707. **[Crossref]**

Acta Medica 2026; Early View: 174-183 DOI: 10.32552/actamedica.2026.1303

acta medica

RESEARCH ARTICLE

Is primary hyperparathyroidism associated with less aggressive histological subtypes and clinicopathological features of papillary thyroid carcinoma? A large single-center cohort study

İbrahim Kılınç[1], Mustafa Oruç[1], Alparslan Ertenlice[1,2]

[1]Department of General Surgery, Ankara Bilkent City Hospital, Ankara, Türkiye
[2]Department of General Surgery, Faculty of Medicine, Ankara Yıldırım Beyazıt University, Ankara, Türkiye

Abstract

Objective: The coexistence of primary hyperparathyroidism (PHPT) and papillary thyroid carci-noma (PTC) has been increasingly recognized; however, its impact on tumor behavior and clinicopathological features remains unclear. To investigate the association be-tween PHPT and the clinicopathological characteristics of PTC in a large single-center cohort.

Materials and Methods: This retrospective study included patients who underwent parathyroidectomy for PHPT between 2019 and 2024. Patients with concomitant PTC were identified and compared with a separate cohort of patients with PTC without PHPT. Demographic, biochemical, and clinicopathological features, including tumor subtypes and adverse pathological characteristics, were analyzed.

Results: Among 190 patients with PTC, 91 (47.9%) had concomitant PHPT. Patients with PHPT were older and more frequently female. Tumor size was significantly smaller in patients with PHPT. Aggressive histological subtypes were significantly less frequent in the PHPT group (8.8% vs. 21.2%, p = 0.017). In addition, capsular invasion, lym-phovascular invasion, perineural invasion, and lymph node metastasis were observed less frequently in patients with PHPT. Radioactive iodine use was also significantly lower in this group.

In the PHPT cohort (n = 750), the presence of concomitant PTC was not associated with significant differences in preoperative calcium, phosphorus, PTH, or ALP levels. However, patients with PTC had lower preoperative magnesium levels and exhibited distinct postoperative biochemical profiles.

Conclusion: PHPT may be associated with non-aggressive subtypes and more favorable clinico-pathological features in PTC. However, these findings should be interpreted cautious-ly, as tumor size and other potential confounders may influence the observed associa-tions. PHPT alone should not be considered a determinant for treatment de-escalation, and clinical decision-making should remain guided by established risk-adapted strategies. These findings may provide additional insight into risk stratification in patients with coexisting PHPT and PTC.

Keywords: hyperparathyroidism, thyroid cancer, papillary

Corresponding author: İbrahim Kılınç • **Email:** ikilinc8083@gmail.com

Received: April 13, 2026 **Accepted:** April 28, 2026 **Published online:** June 28, 2026

Introduction

Primary hyperparathyroidism (PHPT) is a common endocrine disorder affecting ap-proximately 0.1–0.4% of the general population [1]. In the majority of cases, it is caused by a solitary parathyroid adenoma, and surgical excision remains the definitive treatment [2]. Given the relatively high prevalence of con-comitant thyroid pathology, preoperative thyroid evaluation is routinely recommended in patients undergoing parathyroidectomy [3].

Papillary thyroid carcinoma (PTC), accounting for 85–90% of all thyroid malignancies, is the most common type of thyroid cancer, with a steadily increasing incidence worldwide over recent decades [4,5]. Although several stud-ies have reported an association between PHPT and thyroid carcinoma [6,7], the nature of this relationship remains incompletely understood [8]. In particular, it is still unclear whether this coexistence reflects a causal link or represents a coincidental finding, and further investigation is required to clarify the underlying mechanisms.

Potential shared molecular pathways and genetic predispositions contributing to the coexistence of these two conditions have been proposed, but current evidence re-mains limited [9,10]. Moreover, comprehensive studies eval-uating the impact of PHPT on the prognosis, disease progression, and treatment out-comes of PTC—as well as the potential influence of PTC on the clinical course of PHPT—are scarce [11,12].

A better understanding of the clinical implications of this coexistence is of considera-ble importance, particularly in terms of diagnostic challenges, surgical planning, and treatment strategies [7,12-14].

Therefore, the aim of this study was to systematically evaluate the potential interac-tion between PHPT and PTC in a large patient cohort.

Materials and Methods

Study design and patient selection

This retrospective study was conducted in two phases. In the first phase, all consecu-tive patients with PHPT who underwent surgery at our institution between January 2019 and December 2024 were included. All procedures were performed by attending surgeons from the Department of General Surgery.

During this period, a total of 816 patients underwent parathyroidectomy. Of these, 67 patients were excluded due to secondary or tertiary hyperparathyroidism (n = 36), fa-milial syndromes such as MEN1 or MEN2A (n = 7), incomplete medical records (n = 17), or unsuccessful parathyroidectomy (n = 6). After applying these exclusion criteria, 750 patients were included in the final analysis and were stratified according to the presence or absence of concomitant PTC (Figure 1).

In the second phase, to enable comparison between patients with PTC with and with-out PHPT, a separate cohort of patients with PTC without PHPT was included. These patients were selected from consecutive cases who underwent surgery between Jan-uary 2019 and December 2020 and had a histopathological diagnosis of PTC, in order to achieve a comparable sample size with the PTC + PHPT group. This approach was chosen to minimize temporal bias and ensure a homogeneous surgical and diagnostic environment following the establishment of our high-volume center in 2019.

Patients with concomitant PHPT and PTC were included in both analytic cohorts, as each analysis addressed a distinct research question.

Data collection

Preoperative and postoperative biochemical data were retrieved from institutional electronic medical records. All biochemical analyses were performed in the central laboratory of our institution using standardized automated analyzers in accordance with the manufacturers' instructions.

Alkaline phosphatase (ALP), calcium, phosphorus, magnesium, and creatinine levels were measured using the Siemens Atellica CH automated chemistry analyzer (Sie-mens Healthineers, Erlangen, Germany), employing IFCC-recommended kinetic col-orimetric or enzymatic methods, as appropriate. Parathyroid hormone (PTH) levels were measured using chemiluminescent immunoassay on the Siemens Atellica IM analyzer.

The reference ranges were as follows: ALP 42–98 U/L, calcium 8.7–10.4 mg/dL, phosphorus 2.4–5.1 mg/dL, magnesium 1.3–2.7 mg/dL, and creatinine 0.5–1.1 mg/dL. The reference range for serum PTH was 18.4–80.1 ng/mL.

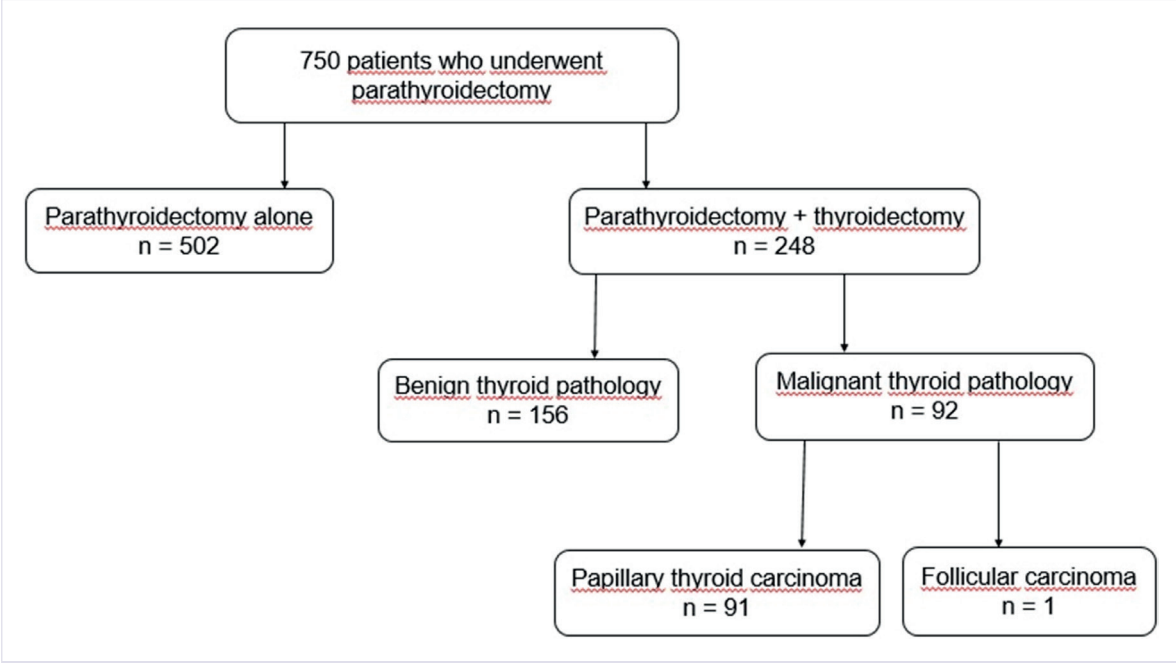


Figure 1. Flowchart of the study population and patient selection process

Serum calcium concentrations were corrected for albumin using the following formula:

Corrected calcium (mg/dL) = [0.08 × (40 − albumin (g/L))] + measured calcium.

Laboratory parameters were collected at predefined perioperative time points. Magne-sium and phosphorus levels obtained during the final week prior to surgery were rec-orded. PTH levels included the highest preoperative value measured within one month before surgery and the postoperative value measured on postoperative day 1. Similar-ly, calcium levels were defined as the highest preoperative value within one month prior to surgery and the postoperative value measured on postoperative day 1.

Clinical and pathological data

Data on PTC were obtained from pathology reports. Parathyroid adenoma localization and dimensions were determined based on preoperative ultrasonography and opera-tive records. Adenomas were assumed to have an ellipsoid shape, and their volumes were calculated using the formula: volume = (π/6) × length × width × depth.

Patients with renal stones detected on imaging (ultrasonography or computed tomog-raphy) or with a documented history of nephrolithiasis were classified as having a his-tory of renal stones.

Definitions

PHPT with PTC was defined as the coexistence of histopathologically confirmed para-thyroid tumor and PTC in the same patient, with both lesions identified and treated during the same surgical procedure.

PTC subtypes were classified according to the World Health Organization criteria.

Aggressive PTC subtypes were defined as tall cell, hobnail, columnar cell, diffuse sclerosing, and solid/trabecular variants, based on their established association with adverse clinicopathological features. Other subtypes—including classic (conventional), follicular, Warthin-like, oncocytic, and cribriform–morular variants—were not classified as aggressive.

Bone health was assessed using bone mineral density (BMD) and T-scores obtained from dual-energy X-ray absorptiometry (DXA) of the lumbar spine. Lumbar spine measurements were selected because they represented the most consistently availa-ble data in this retrospective cohort (missing in only 8 patients). Measurements from the radius, femoral neck, and total

hip were not consistently available and were there-fore excluded from analysis.

DXA measurements were performed using a GE/Lunar Prodigy densitometer (GE Healthcare, Madison, WI, USA) according to standard protocols. The precision error for lumbar spine BMD measurements was within acceptable limits, with a coefficient of variation (CV) ≤ 2%, in accordance with international densitometry standards.

Statistical analysis

Continuous variables were assessed for normality using visual inspection and the Shapiro–Wilk test and are presented as mean ± standard deviation (SD) or median with interquartile range (IQR), as appropriate. Categorical variables are expressed as frequencies and percentages.

Comparisons between groups were performed using Student's t-test or the Mann–Whitney U test for continuous variables and the chi-square test or Fisher's exact test for categorical variables, as appropriate.

A two-sided p-value < 0.05 was considered statistically significant. All statistical anal-yses were performed using R software (version 4.1.1).

Ethical approval

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki. The study protocol was approved by the local Ethics Committee (approval number: TABED 1-25-1568). Informed consent was obtained from all participants as part of the routine institutional admission procedure, which includes permission for the retrospective use of anonymized clinical data for scientific research purposes.

Results

Comparison of PTC with and without concomitant PHPT

Among 190 patients with PTC, 91 (47.9%) had concomitant PHPT. Patients with PTC and PHPT were older than those without PHPT (55 ± 11 vs. 47 ± 14 years, p < 0.001) and were more frequently female (89% vs. 71.7%, p = 0.003). The distribution of con-ventional (classic) PTC was similar in patients with and without PHPT (78.1% vs. 78.7%). In contrast, the follicular subtype was more common in patients without PHPT (42.4% vs. 33.0%) (Figure 2). Aggressive PTC subtypes were more frequently ob-served in patients without PHPT (21.2% vs. 8.8%, p = 0.017).

Patients without PHPT demonstrated higher rates of adverse pathologic features, in-cluding capsule invasion (31.3% vs. 17.6%, p = 0.028), lymphovascular invasion (20.2% vs. 6.6%, p = 0.006), perineural invasion (11.1% vs. 3.3%, p = 0.039), and central lymph node metastasis (29.3% vs. 6.6%, p < 0.001). Lateral lymph node me-tastasis was observed exclusively in patients without parathyroid adenoma (PA) (7.9% vs. 0%, p < 0.001). Radioactive iodine therapy was administered more frequently to patients without PHPT (80.8% vs. 21.1%, p < 0.001). The rates of multifocality, bilat-erality, recurrence, and follow-up duration did not differ significantly between the groups (Table 1).

Comparison of PHPT with and without concomitant PTC

A total of 750 patients with parathyroid pathology were included, of whom 91 (12.1%) had concomitant PTC. Patients with PHPT and concomitant PTC were slightly older than those without PTC (55 ± 11 vs. 52 ± 12 years, p = 0.002) and were more fre-quently female (89% vs. 80%, p = 0.03).

The preoperative biochemical parameters, including serum calcium, phosphorus, par-athyroid hormone (PTH) and ALP levels, were largely comparable between the groups. However, patients with concomitant PTC had significantly lower preoperative magnesium levels (median 1.99 vs. 2.07 mg/dL, p = 0.003).

Postoperatively, patients with PA and PTC had significantly lower calcium levels (me-dian 8.48 vs. 8.90 mg/dL, p < 0.001), higher phosphorus levels (3.50 ± 0.81 vs. 3.21 ± 0.64 mg/dL, p = 0.001), and lower PTH levels (median 10 vs. 17 pg/mL, p = 0.002). Parathyroid gland location, pathology subtype, size, and volume did not differ signifi-cantly between the groups. Similarly, the prevalence of renal stones, bone mineral density, and T-scores were comparable between patients with and without concomi-tant PTC (Table 2).

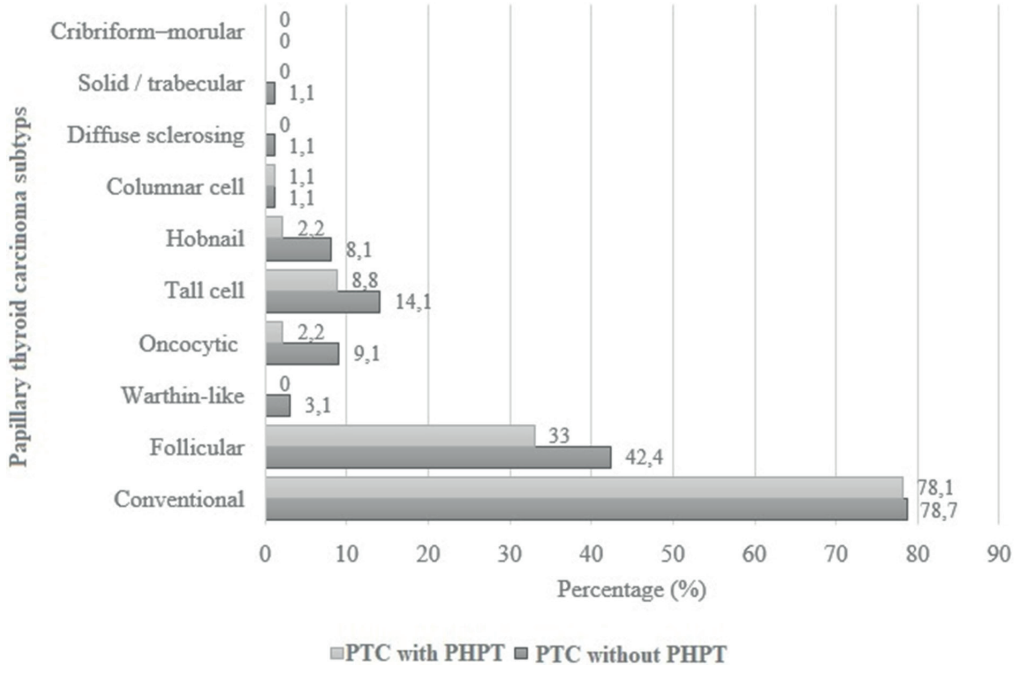


Figure 2. Distribution of PTC subtypes in patients with or without concomitant PHPT

Percentage distribution of papillary thyroid carcinoma subtypes in patients with and without parathyroid adenomas. PHPT: Primary Hyperparathyroidism; PTC: papillary thyroid carcinoma. Percentages may exceed 100% due to the presence of mixed histological patterns.

Table 1. Comparison of PTC with or without PHPT

Variables	Total (n=190)	PTC without PHPT (n=99)	PTC with PHPT (n=91)	p
Age, years, mean, SD	51 ± 14	47 ± 14	55 ± 11	<0.001
Sex, female, n, %	152 (80)	71 (71.7)	81 (89)	0.003
Tumor Size, cm, median, IQR	12 (6-17)	15 (12-25)	6 (3-9)	<0.001
Aggressive tumor variant, n, %	29 (15.3)	21 (21.2)	8 (8.8)	0.017
Multifocal tumor, n, %	86 (45.3)	47 (47.5)	38 (42.9)	0.523
Bilateral tumor, n, %	61 (32.1)	35 (35.4)	26 (28.6)	0.317
Capsule Invasion, n, %	47 (24.7)	31 (31.3)	16 (17.6)	0.028
Extrathyroidal extension, n, %	41 (21.7)	27 (27.3)	14 (15.6)	0.051
Lymphovascular Invasion, n, %	26 (13.7)	20 (20.2)	6 (6.6)	0.006
Perineural Invasion, n, %	14 (7.4)	11 (11.1)	3 (3.3)	0.039
Central Lymph Node Metastasis, n, %	35 (18.4)	29 (29.3)	6 (6.6)	<0.001
Lateral Lymph Node Metastasis, n, %	15 (7.9)	15 (15.1)	0 (0)	<0.001
Radioactive Iodine Treatment, n, %	99 (52.4)	80 (80.8)	19 (21.1)	<0.001
Recurrence, n, %	3 (1.6)	3 (3)	0 (0)	0.247
Follow op time, months, median, IQR	18 (10-35)	16 (7-33)	20 (11-35)	0.09

IQR: Interquartile range, PHPT: Primary Hyperparathyroidism, PTC: Thyroid Papillary Carcinoma, SD: Standard Deviation.

Table 2. Comparison of PHPT with or without PTC

Variables	Total (n=750)	PHPT without PTC (n=659)	PHPT with PTC (n=91)	p
Age, years, mean, SD	53 ± 12	52 ± 12	55 ± 11	0.002
Sex, female, n, %	608 (81.1)	527 (80)	81 (89)	0.03
Preoperative Laboratory Values				
ALP, U/L, median, IQR	108 (85-137)	107 (84-136)	109 (88-149)	0.313
Calcium, mg/dL, median, IQR	11.52 (11.08-11.96)	11.53 (11.09-11.97)	11.41 (11.03-11.92)	0.483
Phosphorus, mg/dL, mean, SD	2.62 ± 0.56	2.61 ± 0.55	2.69 ± 0.62	0.109
Magnesium, mg/dL, median IQR	2.06 (1.91-2.19)	2.07 (1.92-2.2)	1.99 (1.83-2.13)	0.003
Creatinine, mg/dL, median, IQR	0.74 (0.64-0.86)	0.73 (0.64-9.85)	0.76 (0.64-0.87)	0.483
PTH, pg/mL, median, IQR	192 (139-273)	191 (139-271)	205 (137-311)	0.650
Postoperative Laboratory Findings				
Calcium, mg/dL, median, IQR	8.85 (8.44-9.32)	8.90 (8.49-9.37)	8.48 (7.91-8.89)	<0.001
Phosphorus, mg/dL, mean, SD	3.25 ± 0.67	3.21 ± 0.64	3.50 ± 0.81	0.001
PTH, pg/mL, median, IQR	16 (9-35)	17 (9-36)	10 (4.5-29.7)	0.002
Parathyroid location, n, %				0.119
Left Superior	88 (11.7)	81 (12.3)	7 (7.7)	
Left Inferior	273 (36.4)	240 (36.4)	33 (36.3)	
Right Superior	89 (11.9)	82 (12.4)	7 (7.7)	
Right Inferior	248 (33.1)	212 (32.2)	36 (39.6)	
Intrathyroidal	34 (4.5)	31 (4.7)	3 (3.3)	
Atypic	18 (2.4)	13 (2)	5 (5.5)	
Parathyroid pathology n, %				0.567
Adenoma	687 (91.8)	605 (91.9)	82 (90.1)	
Atypic adenoma	25 (3.3)	22 (3.3)	3 (3.3)	
Hyperplasia	32 (4.3)	26 (4)	6 (6.6)	
Cancer	5 (0.7)	5 (0.7)	0 (0)	
Parathyroid size, mm, median, IQR	12 (9-17)	12 (9-16)	12 (9-18)	0.343
Parathyroid volume mm3, median, IQR	374.6 (164.8-918.3)	363 (164.4-863.9)	457 (164-1706)	0.121
History of Renal Stone, n, %	169 (23)	153 (23.8)	16 (17.6)	0.188
BMD g/cm3, median, IQR	0.99 (0.88-1.10)	0.99 (0.88-1.10)	0.96 (0.82-1.13)	0.170
T Score, median, IQR	-1.40 (-2.40 - -0.5)	-1.40 (-2.40 - -0.50)	-1.70 (-2.70 - -0.4)	0.160

ALP: Alkaline Phosphatase, BMD: Bone Mass Density, IQR: Interquartile range, PHPT: Primary Hyperparathyroidism, PTH: Parathyroid Hormone, PTC: Thyroid Papillary Carcinoma, SD: Standard Deviation.

Discussion

In this large cohort study, we compared patients with PTC with and without concomi-tant PHPT in terms of demographic, biochemical, and clinicopathological characteris-tics. Our findings suggest that the presence of PHPT may be associated with differ-ences in tumor behavior and could represent a clinically relevant factor in the evalua-tion of PTC. In addition, we performed a secondary analysis comparing patients with isolated

PHPT and those with concomitant PTC across similar parameters.

Previous studies investigating the coexistence of PHPT and PTC have predominantly focused on the prevalence of thyroid disease in patients with PHPT, frequently report-ing a high incidence of papillary thyroid microcarcinomas (PTMCs) and incidentally detected tumors in this population [14-17]. However, to the best of our knowledge, large-scale studies systematically comparing the distribution of aggressive histological subtypes and other adverse clinicopathological features be-tween PTC patients with and without PHPT remain limited. In the present study, we evaluated not only tumor frequency but also a wide range of prognostic parameters reflecting tumor biology and clinical behavior. Our findings suggest that PTC associat-ed with PHPT may exhibit a more indolent clinicopathological phenotype compared with isolated PTC. Nevertheless, a causal relationship cannot be established based on the current data.

The existing literature on this topic is heterogeneous. Beebeejaun et al. proposed that elevated PTH levels, hypercalcemia, and low 1,25-dihydroxyvitamin D levels may in-fluence the tumor microenvironment and thereby alter the biological behavior of PTC [8]. Similarly, Jeong et al. and Çetin et al. reported that PTC associated with PHPT may present with more aggressive histopathological and clinical features [6,7]. In contrast, Hu et al. found no significant difference in tumor aggressiveness between PTC patients with and without PHPT, while Tsai et al. suggested that PTC associated with PHPT may follow an indolent course, potentially due to earlier detection [18]. These conflicting findings highlight the need for further well-designed studies to clarify the nature of this association.

One of the most notable findings of our study is the significantly lower frequency of aggressive histological subtypes—such as tall cell and hobnail variants, which are well known to be associated with poor prognosis and higher invasive potential—in patients with concomitant PHPT [19]. In addition, rates of capsular invasion, lymphovascular invasion, and perineural invasion were significantly lower in this group, while extrathyroidal extension showed a trend toward lower frequency. The lower rate of lymph node metastasis further supports the notion that these tumors may exhibit a more indolent biological behavior [20-22].

However, the significantly smaller tumor size observed in patients with concomitant PHPT should be considered an important confounding factor that may partially account for the lower rates of invasion and metastasis. Therefore, it would be inappro-priate to attribute the less aggressive clinicopathological features solely to the pres-ence of PHPT without accounting for tumor size differences. In this context, tumor size should be taken into consideration when interpreting the observed differences between groups.Adjuvant treatment patterns may also reflect these differences in tu-mor characteristics. The lower rate of radioactive iodine (RAI) use in patients with concomitant PHPT may suggest that these tumors are more frequently classified into lower-risk categories. Given that current guidelines recommend a more selective use of RAI in low-risk PTC patients [23,24], this finding may indi-cate that unnecessary adjuvant treatments could potentially be avoided in this sub-group. However, the difference in RAI use should be interpreted cautiously, as it likely reflects risk-adapted clinical decision-making rather than intrinsic tumor biology.

It has been suggested that PTC associated with PHPT may be detected at smaller sizes due to increased diagnostic scrutiny, leading to a more indolent clinical course [25]. This raises the possibility that detection bias may contribute to the observed differences in tumor behavior. However, the lower frequency of aggressive histological variants observed in our study suggests that the indolent phenotype can-not be explained solely by earlier detection and may also reflect underlying biological differences.

From a biological perspective, the chronic hypercalcemic microenvironment character-istic of PHPT may modulate calcium-sensing receptor (CaSR)-mediated signaling pathways, thereby influencing cellular proliferation and differentiation processes [26,27]. In addition, the observed indolent phenotype may speculatively reflect differences in the underlying molecular landscape, including a lower prevalence of alterations associated with aggressive behavior, such as BRAFV600E and TERT promoter mutations [28]. However, these hy-potheses require validation in future molecular and prospective studies.

The lower frequency of aggressive histological variants in PTC associated with PHPT may also be considered in the context of the risk-adapted approach emphasized in the American Thyroid Association (ATA) 2025 guidelines. Current recommendations de-fine recurrence risk

based on a combination of pathological features, with the accumu-lation of adverse characteristics leading to higher risk categories [24]. In this regard, the lower rates of aggressive subtypes, lymph node metastasis, and other invasive features observed in patients with PHPT suggest that these tumors may be more likely to fall into lower-risk categories. However, given the differences in tumor size between groups and the absence of multivariable analyses, caution is warranted when translating these findings into clinical decision-making.

In our secondary analysis comparing patients with PHPT with and without concomi-tant PTC, preoperative biochemical parameters were largely similar between groups. The absence of significant differences in serum calcium, phosphorus, PTH, and ALP levels suggests that the presence of PTC is not associated with the biochemical se-verity of PHPT. This finding supports the notion that the two conditions may largely follow independent biochemical courses. Interestingly, preoperative magnesium levels were significantly lower in patients with concomitant PTC. Given the potential role of magnesium in cellular proliferation and oxidative stress, this finding may indicate a possible interaction between mineral metabolism and thyroid tumor biology. Our re-sults are consistent with previous studies reporting an association between low serum magnesium levels and thyroid cancer [29,30]. Postoperative-ly, patients with concomitant PTC exhibited lower calcium and PTH levels and higher phosphorus levels, suggesting a different balance of mineral metabolism following surgery. However, the absence of differences in parathyroid gland size, volume, local-ization, and pathological characteristics suggests that these findings may be more closely related to surgical extent or perioperative factors rather than intrinsic disease burden. Furthermore, the similar rates of renal stone history, bone mineral density, and T-scores between groups indicate that end-organ involvement in PHPT appears to be independent of the presence of concomitant PTC. Taken together, these find-ings suggest that the coexistence of PTC does not substantially alter the systemic clinical manifestations of PHPT.

Several limitations of this study should be acknowledged. First, the retrospec-tive design and baseline differences between groups—particularly in age, sex, and tumor size—limit the ability to fully account for potential confounding factors. Second, the absence of multivariable analyses precludes definitive assessment of the inde-pendent effect of PHPT. Finally, the relatively short follow-up period restricts the eval-uation of recurrence and long-term oncological outcomes.

In conclusion, this study suggests that PHPT may be associated with non-aggressive subtypes of PTC and that these tumors may exhibit less aggressive clinicopathologi-cal features compared with isolated PTC. The lower frequency of aggressive histolog-ical variants, invasive features, and lymph node metastasis supports the notion of a more indolent tumor phenotype. While the more favorable pathological profile ob-served in patients with concomitant PHPT appears to be consistent with the risk-adapted approach emphasized in the ATA 2025 guidelines, PHPT itself should not be considered an independent criterion for treatment de-escalation. Clinical decision-making in this patient population should remain guided by established guidelines and multidisciplinary evaluation. These findings may provide additional insight into risk stratification in patients with coexisting PHPT and PTC.

Ethical approval

This study was approved by the local Ethics Committee (approval number: TABED 1-25-1568).

Author contributions

Conception: İ.K., M.O., A.E.; Design: İ.K.; Data acquisition: İ.K., A.E.; Data analysis: M.O.; Data interpretation: İ.K., M.O., A.E.; Drafting of the manuscript: İ.K., A.E.; Critical revision of the manuscript: İ.K., M.O., A.E. All authors reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of this study.

Ethical approval

This study was approved by the Ankara Bilkent City Hospital-Medical Research and Clinical Ethics Evaluation Committee (Date: 13.08.2025, Decision/Protocol No: TABED 1-25-1568). Informed consent was obtained from all participants involved in this study.

Data availability statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflict of interest

The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The authors declare that this study received no funding.

Generative AI statement

The authors declare that no generative AI or AI-assisted technologies were used in the writing or preparation of this study.

References

[1] Heath DA. Primary hyperparathyroidism. Clinical presentation and factors influencing clinical management. Endocrinol Metab Clin North Am 1989;18(3):631-46.

[2] Fraser WD. Hyperparathyroidism. Lancet 2009;374(9684):145-58. **[Crossref]**

[3] Wilhelm SM, Wang TS, Ruan DT, et al. The American Association of Endocrine Surgeons Guidelines for definitive management of primary hyperparathyroidism. JAMA Surg 2016;151(10):959-68. **[Crossref]**

[4] Noone AM, Howlader N, Krapcho M, et al. SEER cancer statistics review, 1975-2015. National Cancer Institute 2018. Available at: https://seer.cancer.gov/csr/1975_2015/.

[5] Vaccarella S, Dal Maso L, Laversanne M, Bray F, Plummer M, Franceschi S. The impact of diagnostic changes on the rise in thyroid cancer incidence: a population-based study in selected high-resource countries. Thyroid 2015;25(10):1127-36. **[Crossref]**

[6] Çetin K, Sıkar HE, Temizkan Ş, et al. Does primary hyperparathyroidism have an association with thyroid papillary cancer? A retrospective cohort study. World J Surg 2019;43(5):1243-8. **[Crossref]**

[7] Jeong C, Kwon HI, Baek H, et al. Association of hyperparathyroidism and papillary thyroid cancer: a multicenter retrospective study. Endocrinol Metab (Seoul) 2020;35(4):925-32. **[Crossref]**

[8] Beebeejaun M, Chinnasamy E, Wilson P, Sharma A, Beharry N, Bano G. Papillary carcinoma of the thyroid in patients with primary hyperparathyroidism: is there a link? Med Hypotheses 2017;103:100-4. **[Crossref]**

[9] Vargas-Ortega G, Balcázar-Hernández L, González-Virla B, et al. Symptomatic primary hyperparathyroidism as a risk factor for differentiated thyroid cancer. J Thyroid Res 2018;2018:9461079. **[Crossref]**

[10] Emami A, Nazem MR, Shekarriz R, Hedayati M. Micronutrient status (calcium, zinc, vitamins D and E) in patients with medullary thyroid carcinoma: a cross-sectional study. Nutrition 2017;41:86-89. **[Crossref]**

[11] Nilsson IL, Zedenius J, Yin L, Ekbom A. The association between primary hyperparathyroidism and malignancy: nationwide cohort analysis on cancer incidence after parathyroidectomy. Endocr Relat Cancer 2007;14(1):135-40. **[Crossref]**

[12] Palmieri S, Roggero L, Cairoli E, et al. Occurrence of malignant neoplasia in patients with primary hyperparathyroidism. Eur J Intern Med 2017;43:77-82. **[Crossref]**

[13] Chinamon U, Levy D, Marom T. Is primary hyperparathyroidism a risk factor for papillary thyroid cancer? An exemplar study and literature review. Int Arch Otorhinolaryngol 2015; 19: 42-5.

[14] Kutlutürk K, Otan E, Yağcı MA, Usta S, Aydın C, Ünal B. Thyroid pathologies accompanying primary hyperparathyroidism: a high rate of papillary thyroid microcarcinoma. Ulus Cerrahi Derg 2014;30(3):125-8. **[Crossref]**

[15] Li L, Li B, Lv B, et al. Increased thyroid malignancy in patients with primary hyperparathyroidism. Endocr Connect 2021;10(8):885-93. **[Crossref]**

[16] Xue Y, Ye ZQ, Zhou HW, Shi BM, Yi XH, Zhang KQ. Serum calcium and risk of nonmedullary thyroid cancer in patients with primary hyperparathyroidism. Med Sci Monit 2016;22:4482-9. **[Crossref]**

[17] Preda C, Branisteanu D, Armasu I, et al. Coexistent papillary thyroid carcinoma diagnosed in surgically treated patients for primary versus secondary hyperparathyroidism: same incidence, different characteristics. BMC Surg 2019;19(1):94. **[Crossref]**

[18] Tsai CY, Chen ST, Hsueh C, Lin YS, Lin JD. Long-term therapeutic outcomes of papillary thyroid carcinoma with concomitant hyperparathyroidism: a single center case-control study. Biomed J 2020;43(1):53-61. **[Crossref]**

[19] Baloch ZW, Asa SL, Barletta JA, et al. Overview of the 2022 WHO classification of thyroid neoplasms. Endocr Pathol 2022;33(1):27-63. **[Crossref]**

[20] Jeon MJ, Chung MS, Kwon H, et al. Features of papillary thyroid microcarcinoma associated with lateral cervical lymph node metastasis. Clin Endocrinol (Oxf) 2017;86(6):845-51. **[Crossref]**

[21] Jeon MJ, Kim WG, Choi YM, et al. Features predictive of distant metastasis in papillary thyroid microcarcinomas. Thyroid 2016;26(1):161-8. **[Crossref]**

[22] Kim SK, Park I, Woo JW, et al. Predictive factors for lymph node metastasis in papillary thyroid microcarcinoma. Ann Surg Oncol 2016;23(9):2866-73. **[Crossref]**

[23] Haugen BR, Alexander EK, Bible KC, et al. 2015 American Thyroid Association Management Guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association Guidelines task force on thyroid nodules and differentiated thyroid cancer. Thyroid 2016;26(1):1-133. **[Crossref]**

[24] Ringel MD, Sosa JA, Baloch Z, et al. 2025 American Thyroid Association Management Guidelines for adult patients with differentiated thyroid cancer. Thyroid 2025;35(8):841-985. **[Crossref]**

[25] Liu Y, Guo S, Sang S, et al. Differences in clinicopathological characteristics of papillary thyroid carcinoma between symptomatic and asymptomatic patients with primary hyperparathyroidism. Int J Endocrinol 2021;2021:9917694. **[Crossref]**

[26] Saidak Z, Mentaverri R, Brown EM. The role of the calcium-sensing receptor in the development and progression of cancer. Endocr Rev 2009;30(2):178-95. **[Crossref]**

[27] Tennakoon S, Aggarwal A, Kállay E. The calcium-sensing receptor and the hallmarks of cancer. Biochim Biophys Acta 2016;1863(6 Pt B):1398-407. **[Crossref]**

[28] Liu C, Chen T, Liu Z. Associations between BRAF(V600E) and prognostic factors and poor outcomes in papillary thyroid carcinoma: a meta-analysis. World J Surg Oncol 2016;14(1):241. **[Crossref]**

[29] Shen F, Cai WS, Li JL, Feng Z, Cao J, Xu B. The association between serum levels of selenium, copper, and magnesium with thyroid cancer: a meta-analysis. Biol Trace Elem Res 2015;167(2):225-35. **[Crossref]**

[30] Xu H, Hu X, Li J, et al. The Inverse Association of serum magnesium with papillary thyroid cancer in thyroid nodules: a cross-sectional survey based on thyroidectomy population. Biol Trace Elem Res 2023;201(7):3279-89. **[Crossref]**

acta medica

LETTER TO THE EDITOR

Rethinking unresponsiveness: Cognitive motor dissociation in disorders of consciousness

Okan Sökmen[1]

[1]Department of Neurology, Atatürk Sanatoryum Training and Research Hospital, Ankara, Türkiye

Dear Editor,

Disorders of consciousness encompass a range of clinical states, including unresponsive wakefulness syndrome (UWS), minimally conscious states minus and plus (MCS–, MCS+), emergence from MCS (MCS E), and cognitive motor dissociation (CMD) [1,2]. Table 1 summarizes the behavioral, motor, and neurophysiological features of each category. Accurate assessment of consciousness remains one of the most challenging aspects of neurocritical care.

Recent advances in the field of disorders of consciousness have shown that behavior-based clinical assessments may not always accurately reflect a patient's level of consciousness. CMD, also referred to as covert consciousness, is a recently defined concept describing a dissociation between cognitive processes and motor output [3]. This concept challenges established assumptions about consciousness evaluation in patients who appear clinically unresponsive [1,2,4].

Current evidence suggests that patients identified as having CMD differ from other subgroups of disorders of consciousness in terms of brain network organization and clinical characteristics. This indicates that CMD has been described as a potentially distinct subgroup with unique brain network characteristics, although its classification remains to be fully established. In some patients who are clinically diagnosed at the bedside as having UWS or MCS–, the evidence of willful brain activity using advanced neurophysiological methods has led to the recognition of this new entity [2,4]. From a diagnostic perspective, CMD is defined by the detection of voluntary brain responses using task-based functional magnetic resonance imaging (fMRI) or electroencephalography (EEG) in patients who show no behavioral signs of consciousness during bedside examination [2,5,6].

Recent studies have also explored alternative techniques for detecting CMD. For example, a 2025 study using functional near-infrared spectroscopy (fNIRS) combined with a motor imagery paradigm detected covert command-following in a subset of patients who were behaviorally diagnosed as UWS or MCS– [7]. Because fNIRS is portable and can be applied at the bedside, it may represent a more accessible tool for identifying covert consciousness in clinical practice [7,8]. However, its clinical applicability and diagnostic accuracy require further validation. Despite these advances, the routine clinical implementation of these techniques remains limited because of technical complexity, cost, and the need for specialized expertise [8].

It is unclear whether CMD is a stage between MCS– and MCS+, or a separate condition [1,4]. This makes its exact classification within current clinical systems uncertain. Nevertheless, since 2017, studies employing task-based fMRI and EEG paradigms have demonstrated the presence of CMD in a subset of behaviorally unresponsive patients in intensive care units [2,4,5]. The true prevalence of CMD in patients with acute and subacute–chronic disorders of consciousness is not yet

Corresponding author: Okan Sökmen • **Email:** drokansokmen@gmail.com

Received: March 11, 2026 **Accepted:** April 8, 2026 **Published online:** June 28, 2026

Table 1. Clinical comparison of cognitive motor dissociation and other disorders of consciousness [1-6]

Feature	UWS	MCS–	MCS+	MCS-E	CMD
Behavioral response	No observable response	Very limited, inconsistent	Consistent but limited	Functional, reliable	No observable response
Evidence of consciousness	No behavioral evidence	Inconsistent, low-level signs	Clear but limited	Clear	Brain-based evidence only
Brain response to commands	Not detected	Inconsistent	Present	Consistent	Detected (EEG/ fMRI)
Diagnostic method	Bedside examination	Bedside examination	Bedside examination	Bedside examination	Advanced methods (EEG/ fMRI)
Motor output	None	Very limited	Limited but intentional	Functional, intentional	None
Clinical significance/prognosis	Poor	Variable	More Favorable	Favorable	Uncertain

Note: Clinical and neurophysiological findings should be interpreted with caution, as responses may vary and false-negative results may occur depending on patient condition and methodology.
Abbreviations: UWS = Unresponsive Wakefulness Syndrome; MCS– = Minimally Conscious State minus; MCS+ = Minimally Conscious State plus; MCS-E = Emergence from MCS; CMD = Cognitive Motor Dissociation.; EEG: Electroencephalography; fMRI: functional magnetic resonance imaging

known with certainty [1,2]. Reported prevalence rates vary across studies, with some suggesting that CMD may be present in a subset of patients with disorders of consciousness. In addition, CMD detection in the ICU has been associated with functional recovery in some cohorts, although findings remain heterogeneous [1,2]. Recent evidence from a large multicenter prospective cohort study further supports the clinical relevance of cognitive-motor dissociation. In this study, task-based fMRI or EEG detected command-following brain activity in approximately 25% of patients who showed no behavioral response to commands at bedside. These findings highlight that a notable proportion of patients who appear clinically unresponsive may retain covert cognitive processing that is not detectable through behavioral examination [2]. Considering that hundreds of thousands of individuals worldwide live with chronic disorders of consciousness, and that access to advanced diagnostic tools remains limited, the number of unrecognized CMD cases may be higher than currently estimated [2].

At present, meaningful communication with patients diagnosed with CMD is not possible. However, advances in EEG-based brain–computer interface systems raise the possibility that communication with patients in the acute stage of CMD may become feasible in the future [6]. It is still unknown whether a diagnosis of CMD leads to directly actionable treatments, and its prognostic value has not yet been fully clarified. Furthermore, whether and how such findings should be communicated to patients' families raises important ethical concerns [9].

In conclusion, CMD is an important concept that calls for a reconsideration of how disorders of consciousness are evaluated. The integration of advanced neuroimaging and neurophysiological methods into clinical assessment in appropriately selected patients may improve diagnostic accuracy and contribute to more informed prognostic assessment. Importantly, the absence of detectable task-related responses does not exclude preserved consciousness, as false-negative results may occur due to fluctuations in arousal, impaired comprehension, sensory deficits, or methodological limitations. Therefore, negative findings should be interpreted with caution and integrated with clinical and longitudinal assessments. Further studies are needed to clarify the diagnostic, prognostic, and clinical implications of CMD.

Author contribution

Conception and design: O.S.; Data acquisition: O.S.; Data analysis: O.S.; Data interpretation: O.S.; Drafting of the manuscript: O.S.; Critical revision of the manuscript: O.S.; All authors reviewed the results, approved the final version of the manuscript, and agreed to be accountable for all aspects of this study.

Data availability statement

Data sharing is not applicable to this article as no new datasets were generated or analyzed during this study.

Conflict of interest

The authors declares that this study was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Funding

The authors declares that this study received no funding.

Generative AI statement

The authors declares that no generative AI or AI-assisted technologies were used in the writing or preparation of this study.

References

[1] Edlow BL, Claassen J, Schiff ND, Greer DM. Recovery from disorders of consciousness: mechanisms, prognosis and emerging therapies. Nat Rev Neurol 2021;17(3):135-56. **[Crossref]**

[2] Bodien YG, Allanson J, Cardone P, et al. Cognitive motor dissociation in disorders of consciousness. N Engl J Med 2024;391(7):598-608. **[Crossref]**

[3] Schiff ND. Cognitive motor dissociation following severe brain ınjuries. JAMA Neurol 2015;72(12):1413-5. **[Crossref]**

[4] Edlow BL, Chatelle C, Spencer CA, et al. Early detection of consciousness in patients with acute severe traumatic brain injury. Brain 2017;140(9):2399-414. **[Crossref]**

[5] Claassen J, Doyle K, Matory A, et al. Detection of brain activation in unresponsive patients with acute brain injury. N Engl J Med 2019;380(26):2497-505. **[Crossref]**

[6] Chatelle C, Spencer CA, Cash SS, Hochberg LR, Edlow BL. Feasibility of an EEG-based brain-computer interface in the intensive care unit. Clin Neurophysiol 2018;129(8):1519-25. **[Crossref]**

[7] Wang Y, Zeng W, Zou L, et al. Detecting cognitive motor dissociation by functional near-infrared spectroscopy. Front Neurol 2025;16:1532804. **[Crossref]**

[8] Kazazian K, Monti MM, Owen AM. Functional neuroimaging in disorders of consciousness: towards clinical implementation. Brain 2025;148(7):2283-98. **[Crossref]**

[9] Young MJ, Bodien YG, Giacino JT, et al. The neuroethics of disorders of consciousness: a brief history of evolving ideas. Brain 2021;144(11):3291-310. **[Crossref]**